USING CHEMICAL SYNTHESIS TO INVESTIGATE PROTEIN GLYCOSYLATION: O-MANNOSYLATION SITE SPECIFICALLY MODULATES THE STABILITY AND CELLULOSE BINDING AFFINITY OF FAMILY 1 CARBOHYDRATE BINDING MODULES

By

Matthew Robert Drake

B.S., Regis University, 2011

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

Of the requirement for the degree of

Master of Science

2014

This thesis entitled:
Using Chemical Synthesis to Investigate Protein Glycosylation: O-Mannosylation Site Specifically Modulates the Stability and Cellulose Binding Affinity of Family 1 Carbohydrate Binding Modules
written by Matthew Robert Drake
has been approved for the Department of Chemistry and Biochemistry

_____
Zhongping Tan

_____
Richard K. Shoemaker

_____
Tarek Sammakia

Date_____

The final copy of this thesis has been examined by the signatories, and we
Find that both the content and form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

Drake, Matthew Robert (M.S., Chemistry)

Using Chemical Synthesis to Investigate Protein Glycosylation: O-Mannosylation Site Specifically Modulates the Stability and Cellulose Binding Affinity of Family 1 Carbohydrate Binding Modules

Thesis directed by Assistant Professor Zhongping Tan

Lignocellulosic biomass is a massive, but largely unexploited potential source of biofuel. The underutilization of this resource stems largely from the fact that cellulose is difficult to digest into smaller, useable sugar units. Natural lignocellulosic biomass is primarily degraded by fungi, which use Family 1 carbohydrate-binding modules (CBMs) to target cellulose for degradation. Family 1 CBMs are glycosylated, but the effects of glycosylation on CBM function remain unknown. Here, the effects of *O*-mannosylation of the Family 1 CBM from the *Trichoderma reesei* Family 7 cellobiohydrolase (*Tr*Cel7A) are investigated at three glycosylation sites, Thr-1, Ser-3 and Ser-14. The work was made possible by the development of a convenient one-pot synthetic procedure for glycosylated Family 1 CBMs. A library of 20 CBM glycoforms was synthesized with mono-, di-, or tri-saccharides at each glycosylation site. The binding affinity, proteolytic stability, and thermostability of each synthetic glycoform was systematically studied. The results show that even though CBM mannosylation does not induce significant changes to the protein's secondary structure, it can increase the thermolysin cleavage resistance up to 50-fold. Fungi are known to excrete several proteases along with CBM-bearing cellulases, so improved proteolytic stability may improve cellulose digestion efficiency by reducing CBM

degradation. *O*-mannosylation was also shown to increases the thermostability of CBM glycoforms up to 16°C, and a mannose disaccharide at Ser3 has the largest themostabilizing effect. Thermostability is an important property of industrial enzymes because bioreactors are often operated at elevated temperatures. In the binding affinity tests, the glycoforms with small glycans at each site displayed the highest binding affinities for crystalline cellulose, and the glycoform with a single mannose at each of the three positions had the highest binding affinity; a 7.4 fold increase compared to the unglycosyalted CBM. High CBM binding affinity has been linked to increased cellulose digestion rate by fungal cellulases, so these results may have important implications in biofuels production. This study demonstrated how chemical synthesis can be used to systematically study glycosylation and lead to the identification of two CBM glycoforms with particularly desirable stability and binding properties.

I dedicate this work to my wife and parents, for their ongoing love and support.

ACKNOWLEDGMENTS

I would like to thank my coworkers, Dr. Liqun Chen, Erick R. Greene and Patrick K. Chaffey for their work on this project. This project was a true group effort, and it would not have been possible without their contributions. Dr. Chen developed the CBM synthesis and synthesized several of the glycoforms, Erick Green helped with the BMCC binding study and developed automated data analysis for that study and Patrick Chaffey helped synthesize the glycosylated amino acids. I thank Dr. Xiaoyang Guan for teaching me so much about chemical synthesis and laboratory technique. His training made all of my work possible. I thank Professor Zhongping Tan and Dr. Gregg T. Beckham for their guidance and assistance. I thank Hugh O'Neil from the Biofuels Science Focus Area at Oak Ridge National Laboratory for the bacterial cellulose used in this study. I would also like to thank Professor Richard Shoemaker for offering me the job that supported my graduate studies. I wouldn't have stayed to complete my degree without the support. Finally, I want to thank the University of Colorado, Boulder and the US Department of Energy BioEnergy Technologies Office for their financial support of this study.

CONTENTS

TABLES

FIGURES

CHAPTER 1

INTRODUCTION

1.I. Chemical Synthesis as a Tool to Study Glycosylation

Most proteins undergo some form of covalent post-translational modification. These modifications greatly increase the diversity and functional repertoire of proteins. Glycosylation, the attachment of sugars through covalent linkages, is one of the most common post-translational modifications, and at least one-half of human proteins are glycosylated.[1] Compared to other post-translational modifications such as phosphorylation, glycosylation is not very well understood, but it is thought to affect many properties of proteins including: biological activity, solubility, thermostability, and susceptibility to degradation and aggregation.[2]

Glycosylation is not under direct genetic control and glycosylation patterns are highly dependent on local cellular conditions.[3] The result is that different glycosylated forms (glycoforms) of the same protein are often observed within the same organism. It is possible that only one of the many glycoforms of a protein has the desired biological function.[4] As a result, it can be difficult to decipher the effects of glycosylation by studying complex biological mixtures of glycoforms. Abnormal glycosylation states of several proteins have been shown to be indicative of disease,[3] so the ability to study individual glycoforms could prove to be invaluable.

Unfortunately, it is difficult to separate homogeneous glycoforms from biological mixtures and, as a result, it is often only possible to study mixtures of glycoforms using biological expression. Many such studies have been conducted, but they usually only provide an estimate of the average properties and activities of the glycoforms in a complex mixture. Due to

the constant variations in the composition of glycoform mixtures, the results of such studies are often fragmented or inconsistent.[2,5] The ability to isolate individual glycoforms for study would alleviate these problems and allow for the systematic study of glycosylation. Fortunately, chemical synthesis has emerged as a powerful approach to prepare homogeneous peptide and protein glycoforms.[6–8] The synthesis of glycoproteins is more expensive and time consuming than biological expression, but it allows for precise control of glycan structure and makes the incorporation of glycans at unnatural glycosylation sites relatively simple. In principle, chemical synthesis could allow access to any of the glycosylated forms of a protein, which, in turn, would allow the detailed study of the properties of any glycoform.[9,10]

1.II. Biofuels Production and Lignocellulose-Degrading Enzymes

Cellulose is the most abundant, renewable source of reduced carbon in the biosphere. It has been targeted for use in biofuel production, but the difficulties associated with breaking down cellulose crystals into usable fuel sources have made biofuel production from cellulosic biomass relatively inefficient.[11] For this reason, the development of biological systems for the conversion of cellulose into economical ethanol based fuels has attracted significant interest.[11,12]

Most of the plant biomass on our planet is degraded by fungi and bacteria, which secrete synergistic cocktails of enzymes that work in concert to degrade polysaccharides.[11,13] The mono and di-saccharide products of these enzymes can be converted into ethanol for biofuels quite easily by other microorganisms, but the initial degradation of cellulose polymers is usually quite slow, creating a bottleneck in biofuels production. As a result, the study of lignocellulose-

degrading enzymes has become a priority so that ways to increase their efficiency can be identified.

The lignocellulose-degrading enzymes of bacteria and fungi are often multi-modular, consisting of one or more catalytic domains[11,13–17] linked to a carbohydrate-binding module (CBM) that targets plant cell wall polysaccharides through various means.[18] Carbohydrate-binding modules are quite common in nature. To date, 67 CBM families have been identified[19] and many of them are thought to play important roles in the degradation of biomass. The majority of fungal lignocellulose-degrading enzymes that have CBMs have Family 1 CBMs[19] which are small protein domains consisting of less than 40 amino acids. Kraulis *et al*. solved the first Family 1 CBM NMR structure of the well-characterized glycoside hydrolase (GH) Family 7 cellobiohydrolase from the fungus *Trichoderma reesei* (*Hypocrea jecorina*), or *Tr*Cel7A.[20] The structure of the *Tr*Cel7A CBM is a β-sheet rich structure with two disulfide bridges and a flat face formed by aromatic and polar side chains that is believed to be responsible for cellulose binding (Figure 1.1).[20–23]

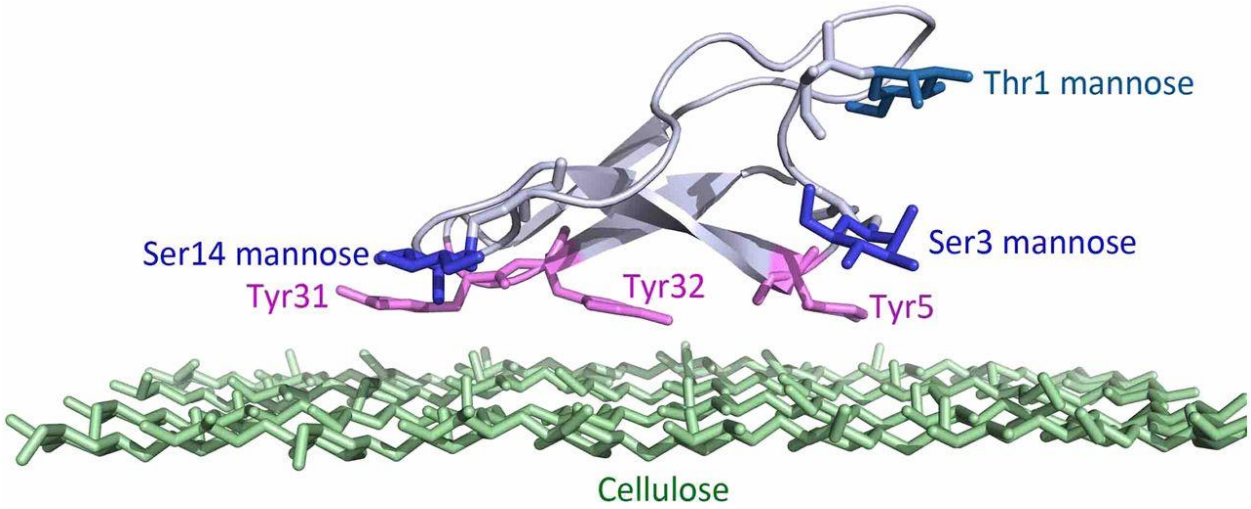

Figure 1.1. The NMR structure of the Family 1 CBM and the top layer of cellulose (11)
The hydrophobic binding face is formed by the three tyrosine residues shown in purple. The three potential glycosylation sites examined in this study are displayed with attached mannoses shown in blue.

1.III. Glycosylation of Lignocellulose-Degrading Enzymes

Fungal enzymes that degrade biomass are often heavily glycosylated.[24,25] Unfortunately, few studies have been conducted on glycosylation in secreted fungal enzymes. In most cases, the extent of lignocellulose-degrading enzyme glycosylation and the factors that control it are unknown. Growth conditions and extracellular glycan-trimming enzymes alter glycosylation patterns so these factors must be carefully controlled if detailed studies are to be performed.

All of the domains of *Tr*Cel7A have been observed to be glycosylated in nature. Catalytic domains have been observed with both *N*- and *O*-linked glycans,[26,27] but the linkers connecting enzymatic domains to CBMs have only been observed to have *O*-linked glycosylation. Linker glycosylation is thought to provide protease protection[28] and may possibly be involved in substrate binding.[29] For *Tr*Cel7A, Harrison *et al.* published the original characterization of the glycosylation pattern on the *Tr*Cel7A linker.[30] The last five residues

analyzed in their study (TQSHY) form the *N*-terminus of the CBM, and the threonine and serine residues (Thr1, Ser3, respectively) were shown to both natively exhibit mannosylation.[30] Since these residues are highly conserved, the observed mannosylation is probably common in Family 1 CBMs.[31] It is also possible that the highly conserved Ser14 residue is naturally mannosylated, but this has not been experimentally characterized. Free energy calculations predict that the mannosylation at Thr-1, Ser-3 and Ser-14 will improve the CBM binding affinity to crystalline cellulose.[32] Furthermore, these calculations suggest that the glycan structure, location, and the number of occupied glycosylation sites will impact the affinity of CBMs for crystalline cellulose.[32] These computational results justified the systematic study of *Tr*Cel7A CBM glycosylation.

To this end, a series of synthetic glycoforms with systematic alterations in glycosylation patterns were produced using chemical glycoprotein synthesis. This small library of CBM glycoforms allowed for the systematic study of the effects of glycosylation on CBM stability and function. The library was constructed so that the effect of glycosylation at each site could be studied independently and in the presence of glycans at other sites to determine the effect of glycoslation at each site and to explore the possibility of synergistic effects from glycosylation at multiple sites. The results of this study should greatly augment the existing knowledge of a protein domain that plays an important role in biomass degradation, and, potentially, biofuels production.

CHAPTER 2

CHEMICAL SYNTHESIS

2.I. Introduction

In order to efficiently produce the library of synthetic CBM glycoforms used in this study, a facile, one-pot synthetic method was developed. The method, which was developed by Dr. Liquin Chen,[33,34] utilized 9-Fluorenylmethoxycarbonyl (Fmoc)-based solid phase peptide synthesis (SPPS) and mannosylated Fmoc amino acid building blocks. Once optimized, the method allowed for each CBM glycoform to be produced in high purity in less than a week. Multiple glycoforms could be synthesized simultaneously and one researcher could reasonably produce up to three a week.

The synthetic approach described herein demonstrates how chemical synthesis can be used to rapidly generate libraries of homogenous small protein glycoforms for comparison studies. Depending on the size and sequence of the protein and the size and complexity of the glycans, the synthetic procedure may take a considerable amount of time to optimize, but once it is the preparation of the glycoforms is fairly facile. For instance, our CBM synthesis took several months to optimize, but after optimization, an individual glycoform could be produced by an undergraduate chemist in about a week. The synthesis optimization time is expected to be reduced as scientists gain more experience and knowledge of glycoprotein synthesis, so this sort of synthetic approach can be expected to become even more efficient in years to come.

2.II. A One-Pot Method for CBM Synthesis and Folding

2.II.a. Solid Phase Peptide Synthesis

Fmoc-based SPPS was used in the synthesis of the CBM glycoforms because of its compatibility with acid-sensitive glycosidic bonds.[35] The glycoforms were synthesized using preloaded trityl resin (Fmoc-Leu-NovaSyn® TGT), and the mannosylated Fmoc amino acid building blocks (Fig. 2).[36] Extended coupling times were used for the mannosylated amino acids to ensure complete coupling.

The initial SPPS resulted in incomplete coupling at Val-18, but a pseudoproline dipeptide strategy[36] using Fmoc-Ala-Ser(psiMe,Mepro)-OH (shown in Figure 2.1 on the following page) was used to solve the problem. The dipeptide presumably induced a conformational change in the growing peptide chain that allowed for complete coupling at Val-18. After SPPS, the peptide was cleaved from the resin and side-chain protecting groups were removed by stirring in 95:2.5:2.5 TFA:TIS:H_2O. The unprotected peptides were then lyophilized and used in the mannose deprotection and folding reactions without purification.

2.II.b. Mannose Deprotection and Folding

The acetyl protecting groups on the mannose residues were removed in 30 min by stirring in 5% hydrazine. The deprotected glycoforms were folded by direct dilution in a mixed glutathione-folding buffer and stirring overnight.[37] The folding products were purified with reverse phase high performance liquid chromatography (HPLC) and lyophilized. The isolated yields varied from 30% for the unglycosylated CBM to 3% for the CBM with three attached mannose trisaccharides. In general, more heavily glycosylated CBM glycoforms had lower synthetic yields.

2.II.c. CBM Glycoform Library

The procedure described above was used to synthesize a total of 20 CBM glycoforms for systematic study. The structures of the glycoforms are illustrated in Figure 2.1.

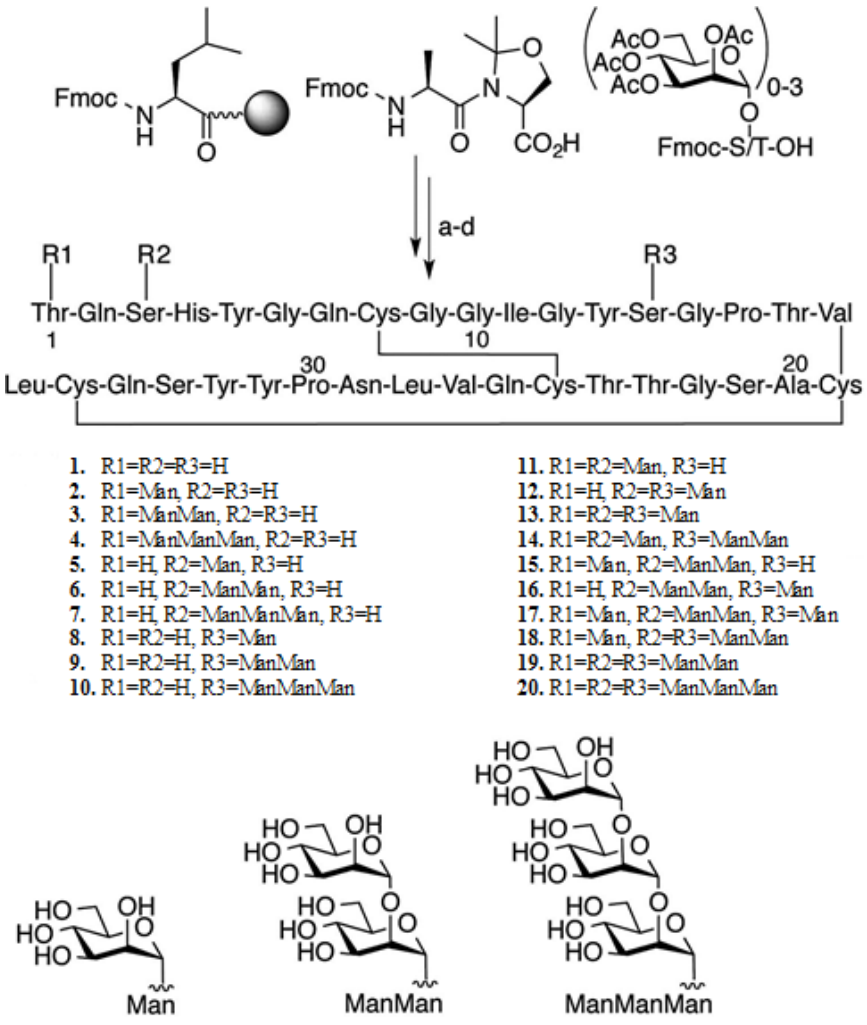

Figure 2.1. The library of CBM glycoforms synthesized for this study. The amino acid building blocks, glycosylation sites, glycan structures and disulfide linkages are also shown.

The library contains the unglycosylated protein, a systematic series of mono-glycosylated analogs (glycoforms 1 through 10 in Figure 2.4) and a series of multiply glycosylated analogs. The mono-glycosyalted analogs allowed for the assessment of the site-specific effects of

glycosylation and the multiply glycosylated analogs allowed for the analysis possible additive or synergistic effects of multiple glycosylation. Together, the library of synthetic CBM glycoforms allowed for the systematic study of CBM glycosylation, which would not be possible using biological mixtures.

2.II.d. Product Verification

In order to confirm the disulfide linkages, the unglycosylated CBM was digested with thermolysin and the resulting fragments were analyzed using LCMS. As shown in Figure 2.5, the peptide fragments that were observed for the digest were consistent with the appropriate disulfide bond pattern (C8 to C25 and C19 to C35). Importantly, the peptide fragments that would have resulted from incorrect disulfide bond connectivity were not observed. Therefore, it is reasonable to conclude that the disulfide connectivity of the unglycosylated CBM product was consistent with the natural protein. Since the *Tr*Cel7A Family 1 CBM is a small protein domain with 2 disulfide bonds, correct disulfide connectivity strongly suggests that the molecule folded correctly.

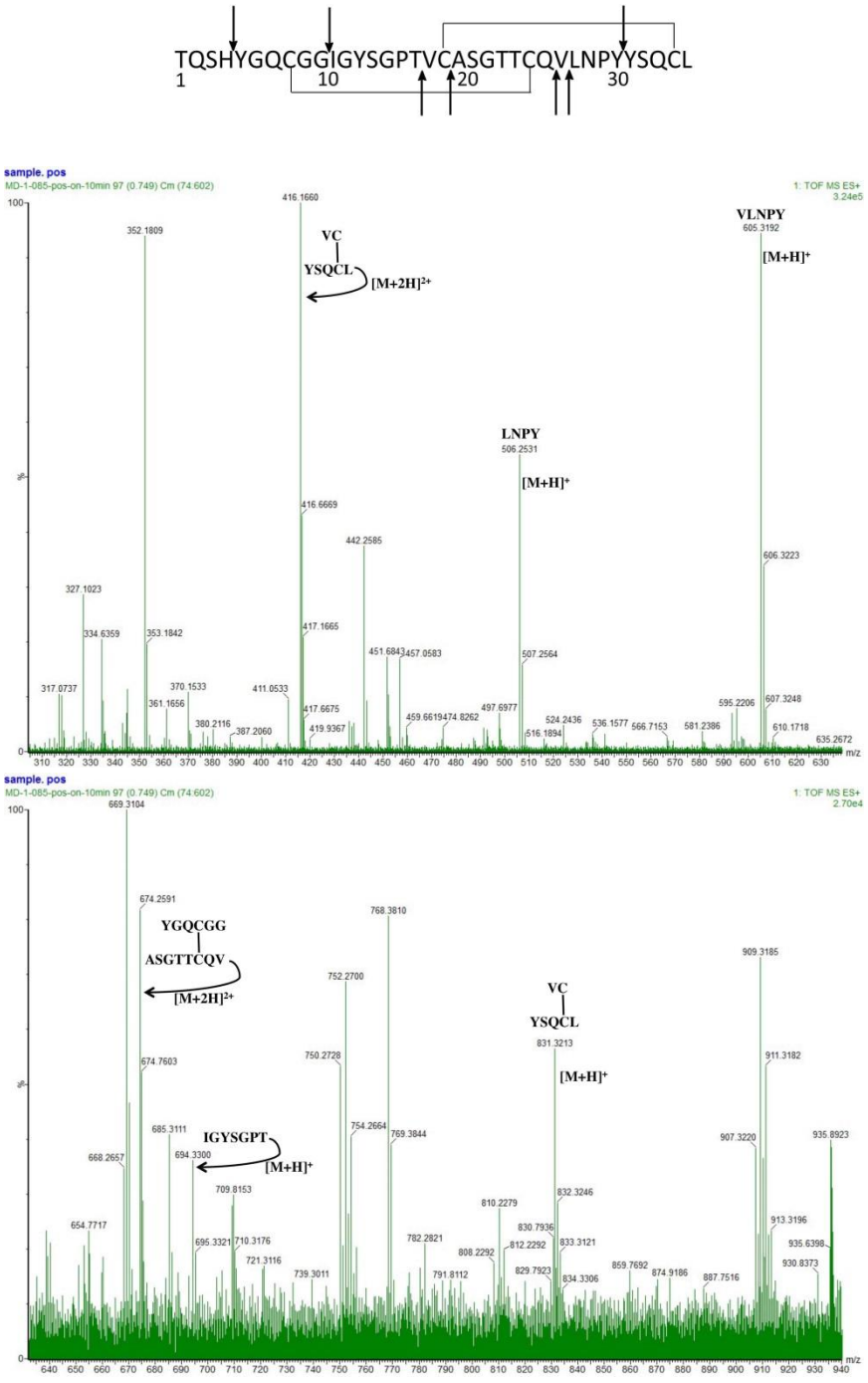

Figure 2.2. Determination CBM disulfide binding by thermolysin digestion. Cleavage sites are indicated by arrows and key fragment peaks are indicated in the mass spectra.

Glycoform identity and purity were experimentally verified by liquid chromatography–mass spectrometry (LC-MS). The LC-MS traces of the purified glycoforms showed a single major peak with the mass of the folded CBM anlogs. LC-MS can distinguish between differently folded proteins with the same mass, so the presence of a single peak in the LCMS trace further provides evidence that the products contained only the correctly folded desired glycoforms. A representative LC-MS trace and ESI mass spectra (for CBM 2) is shown here in Figure 2.6. The LC-MS traces and ESI mass spectra for every CBM glycoform are available in Appendix 3.

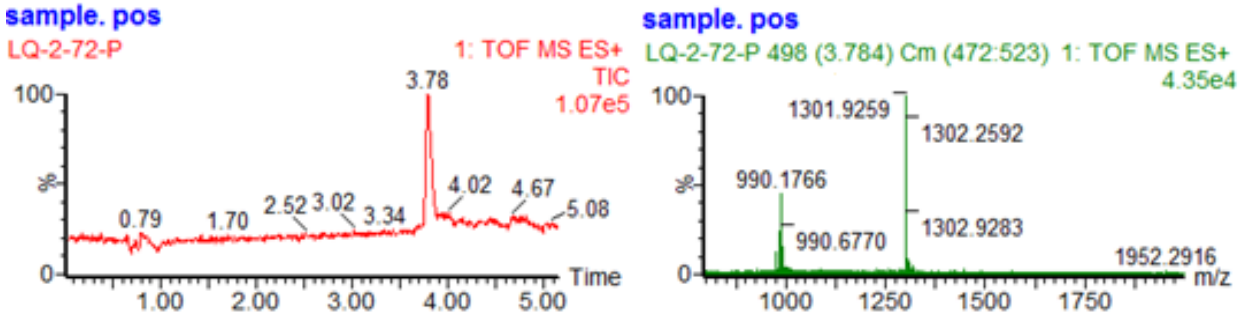

Figure 2.3. The LC-MS trace and ESI-MS of a representative CBM glycoform (CBM 2). MS (ESI) Calcd for $C_{165}H_{245}N_{43}O_{59}S_4$: $[M+2H]^{2+}$ m/z = 1951.32, $[M+3H]^{3+}$ m/z = 1301.21, $[M+4H]^{4+}$ m/z = 976.16.

Next, the secondary structure of the CBM analogs was analyzed using far-UV circular dichroism (CD) spectroscopy. Far-UV CD provides valuable information on the secondary structure of peptides and proteins. The CD spectrum of the unglycosylated CBM was very similar to the spectrum of the unglycosylated molecule presented in literature,[38] which further supports the conclusion that the analog was properly folded. The spectra of most of the analogs are quite similar. They have a large flat depression centered at about 217 nm and a positive peak at about 205 nm, which are consistent with a peptide/protein that adopts predominantly β-sheet

secondary structure (as the CBM does). These results indicate that glycosylation does not cause major secondary structure changes in most of the analogs, which isn't particularly surprising because all three glycosylation sites are in loop or turn regions which are on the exterior of the small peptide. Importantly, O-mannosylation at Thr-1, Ser-3, and Ser-14 sites does not appear to impair CBM folding. The CD spectra of the 20 CBM glycoforms synthesized in this study are provided in Figure 2.7 below. They are fairly similar, with the primary difference being a lower peak at ~204 nm for some glycoforms, suggesting a larger fraction of unordered structure for those glycoforms. The increase in unordered structure could be investigated with H/D exchange.

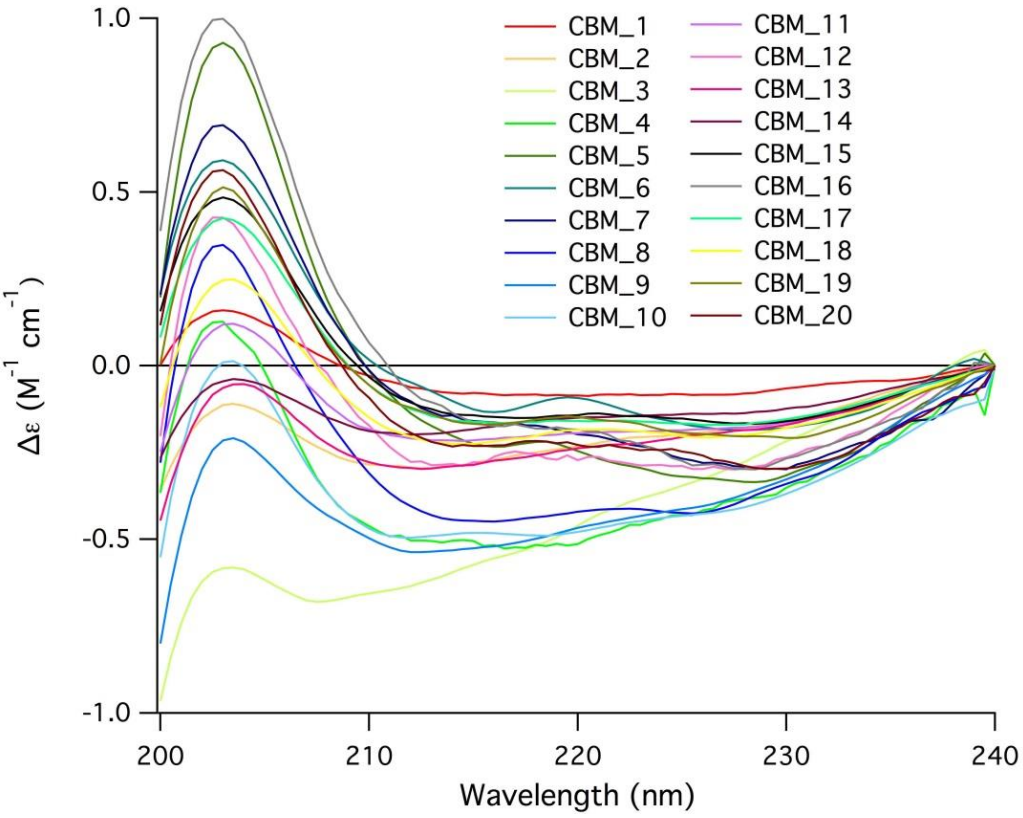

Figure 2.4. The CD spectra of the 20 *Tr*Cel7A CBM glycoforms. All spectra were acquired in 10 mM NaOAc buffer with a pH = 5.2 at 20 °C.

The resulting CD spectra were used to calculate the secondary structure fractions of each CBM analog using the CDPro software provided by Colorado State University. The secondary structure fractions of each CBM analog are fairly similar, supporting the notion that glycosylation at Thr-1, Ser-3 and Ser-14 only slightly changes the conformation. The secondary structure fractions are shown in Table 2.1 below. It does appear that glycosylation slightly decreases the amount of β-sheet structure and increases the amount of "unordered" structure, with more heavily glycosylated analogs generally having less β-sheet structure. Therefore, it can be concluded that CBM glycosylation does not prevent folding but it does slightly disrupt the β-sheet structure of the protein.

Table 2.1. The secondary structure percentages of each CBM glycoform. Calculated from the CD spectra shown in Figure 2.4.

	Secondary Structure Percent			
CBM Variant	B-Sheet	α-helix	Turn	Unordered
1	43.4	2.4	22.9	29.3
2	34.9	3.0	24.0	36.4
3	30.9	1.7	25.6	41.8
4	30.9	0.7	25.2	42.6
5	27.3	0.5	25.1	46.3
6	35.1	0.2	23.5	38.2
7	33.7	0.6	25.1	39.6
8	30.3	1.3	25.7	42.5
9	28.6	1.4	26.1	43.4
10	28.6	0.9	25.9	44.0
11	33.5	~0	23.3	37.0
12	30.6	1.3	24.9	42.5
13	34.2	2.9	23.0	37.6
14	32.3	0.1	25.8	36.2
15	35.3	0.5	23.3	37.7
16	32.3	~0	25.5	42.3
17	34.1	1.6	24.9	38.4
18	34.1	1.6	25.1	38.2
19	34.2	0.7	24.5	40.1
20	34.6	0.8	24.7	39.4

2.II.e. Materials and Methods

All commercial reagents and solvents were used as received. Unless otherwise noted, all reactions and purifications were performed under air atmosphere at room temperature. All LC-MS analyses were performed using a Waters Acquity^{TM} Ultra Performance LC system equipped with Acquity UPLC® BEH 300 C4, 1.7μm, 2.1 x 100 mm column at flow rates of 0.3 and 0.5 mL/min. The mobile phase for LC-MS analysis was a mixture of H_2O (0.1% formic acid, v/v) and acetonitrile (0.1% formic acid, v/v). All preparative separations were performed using a LabAlliance HPLC solvent delivery system equipped with a Rainin UV-1 detector and a Varian Microsorb 100-5, C18 250x21.4mm column at a flow rate of 16.0 mL/min. The mobile phase for HPLC purification was a mixture of H_2O (0.05% TFA, v/v) and acetonitrile (0.04% TFA, v/v). A Waters SYNAPT G2-S system was used mass spectrometric analysis.

Solid-phase peptide synthesis was performed on a PioneerTM Peptide Synthesis System. Peptides and glycopeptides were synthesized by Fmoc chemistry on a pre-loaded Fmoc-Leu-Novasyn® TGT resin. The following Fmoc amino acid building blocks and pseudoproline dipeptides from Chem-Impex, EMD Millipore, and AAPPTec were used in the synthesis: Fmoc-Asn(Trt)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Pro-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Val-OH, and Fmoc-Ala-Ser(psiMe,MePro)-OH. Synthetic cycles were completed with a standard coupling time of 15 min using Fmoc protected amino acids (4 eq.), 2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uroniumhexafluorophosphatemethan aminium (4 eq.) and N,N-diisopropylethylamine (8 eq), except for a prolonged coupling time of 2 h for glycoamino acids. The deblocking was performed by mixing with DMF:piperidine:1,8-

Diazabicyclo[5.4.0]undec-7-ene (100:2:2, v/v/v) for 5 min. Upon completion, the resin was washed into a peptide cleavage vessel with dichloromethane. Cleavage and side-chain deprotection was performed by treatment with TFA:H2O:triisopropylsilane (95:2.5:2.5, v/v/v) solution for 45 min. The filtered cleavage mixture was then concentrated using a gentle stream of air and precipitated at 0 °C by the addition of cold diethyl ether. After centrifugation, the resulting pellet was dissolved in H2O:acetonitrile (1:1, v/v) and lyophilized to dryness for further use.

The acetyl groups of the mannose resiudes were removed by stirring the unpurified synthetic glycopeptides in a hydrazine solution (hydrazine:H2O, 5:100, v/v) at room temperature for 30 min under helium atmosphere. The final concentration of the glycopeptides was 4 mM. The reaction was quenched with a solution of AcOH (AcOH:H2O, 5:100, v/v) and the pH was adjusted to ~8. The folding was initiated by diluting the unprotected glycoforms to a final concentration of ~0.05 mM in a folding buffer (0.2 M Tris-acetate, 0.33 mM oxidized glutathione, 2.6 mM reduced glutathione, pH 8.2). The folding solution was stirred at room temperature for 12 h under helium atmosphere. The solution was then concentrated to a small volume (3-5 mL) using 3 kDa cut-off centrifugal filter units (Amicon) before RP-HPLC purification. The HPLC purification was performed on a Versagrad Preparation-HPLC system using a semi-preparative C-18 column. The products were detected by UV absorption at 275 nm.

The details of the Thermolysin digestion are as follows: Lyophilized Thermolysin, from *Bacillus thermoproteolyticusrokko*, was purchased from Promega Corporation. The digestion was performed in a 50 mM Tris-HCl buffer (pH = 8) with 0.5 mM $CaCl_2$ at a temperature of 37 °C. The reaction was performed in 100 µL of solution with an initial concentration of 270 µM. The solution was prepared so that CBM and thermolysin were initially present in a 20:1 molar

ratio. The reaction was monitored over time by taking 10 µL aliquots and quenching them with an equal volume of 5% AcOH. The aliquots were analyzed using the Waters Acquity UPLC and a Waters SYNAPT G2-S mass spectrometer.

All CD spectra were acquired using an Applied Photophysics Chirascan^{TM}-plus CD spectrometer. In all cases, the spectra were acquired in a 0.5 mm quartz cuvette under nitrogen at a flow rate of 1 L/min. Each CBM analog was dissolved in 10 mM NaOAc buffer with a pH of 5.2. The peptide concentration was 0.2 mg/mL in all tests. CD spectra were obtained at 20 °C with a step of 0.5 nm, 0.5 s per point and a spectral width of 200-240 nm. The spectra are the average of 4 scans with an averaged 4 scan buffer baseline subtracted. The resulting CD spectra were used to calculate the secondary structure fractions of each CBM analog using the CDPro software provided by Colorado State University. The secondary structure fractions of each CBM analog are the average of the results of the three CDPro programs (SELCON3, CDSSTR and CONTIN).

CHAPTER 3

STABILITY STUDIES

3.I. Introduction

While performing the thermolysin digests to determine disulfide bond connectivity, it was observed that the digestion of some analogs proceeded much more rapidly than others. Glycosylation is known to increase the proteolytic stability of some proteins,[2] so it seemed reasonable to investigate if it does so with the CBM glycoforms. Improving the proteolytic stability of cellulases, and their CBMs, may be beneficial to biofuel production. The secretome of *T. reesei* includes a host of proteolytic enzymes, some of which are active under neutral conditions.[39–43] Although these proteolytic enzymes are favorable for biomass conversion since they degrade structural proteins of plants, they also degrade cellulases, which can slow cellulose degradation.[39–44] Glycosylation may improve the proteolytic stability of the CBM (and the cellulase as a whole), which could lead to more efficient cellulose degradation in biofuels production. Thermolysin digests were performed for each CBM analog and the disappearance of the molecular ion peaks were tracked over time. These tests made it possible to determine if glycosylation at certain sites improved the proteolytic stability of the CBM glycoforms.

Cellulase thermostability is important because performing biofuel fermentations at high temperatures increases reaction rates and protects mixtures from microbial contamination.[12,45,46] Recently, the CBM of Cel7A was shown to increase the thermostability of the Cel7A enzyme, and chimeric cellulases with CBMs have also acquired positive thermostabilizing effects.[47,48] Based on these results, it is reasonable to theorize that increasing the thermostability of the CBM

could improve the thermostability of the cellulase. Glycosylation has been shown to increase the thermostability of some proteins,[49] so the thermostability of the 20 CBM glycofoms was investigated using variable temperature circular dichroism to determine if CBM glycosylation has a stabilizing effect.

3.II. Proteolytic Stability

3.II.a. Experimental Design

The CBM glycoforms were digested by thermolysin at 37 °C. Samples were taken at set time intervals for mass analysis. The digestion rate was determined by monitoring the loss of intensity of the molecular ion peak over time using electrospray ionization (ESI) mass spectrometry. Unfortunately, the initial studies yielded inconsistent results, possibly due to the sensitivity of ESI to the salt ions in the digestion buffer. In order to solve this problem, matrix assisted laser desorption/ionization (MALDI) mass spectrometry, which is less sensitive to salt ions, was used to monitor the disappearance of the CBM analogs during digestion. MALDI is not normally thought of as a quantitative mass spectrometry technique, but, with the use of appropriate internal standards, it has been shown the technique can provide reliable quantitative results.[50,51] An appropriate internal standard is structurally similar to the molecule of interest so that it will have similar ionization properties, but it must have a different mass to be observed as a separate peak. Fortunately, the small library of CBM glycoforms could be used as internal standards for each other. Concentration calibration curves were produced over the concentration ranges that were monitored during digestion studies. The mass spectra of the digest samples

were compared to the calibration curves to determine the change in intact CBM glycoform concentration over time. The technique proved to be quite robust, and it was able to monitor CBM glycoform concentrations down to below 1 μM. The data from the digestions were fit to exponential decay functions to determine the half-life of each glycoform in thermolysin. Figure 3.1 shows the data and exponential decay fit for the digest study of the unglycosylated CBM as an example.

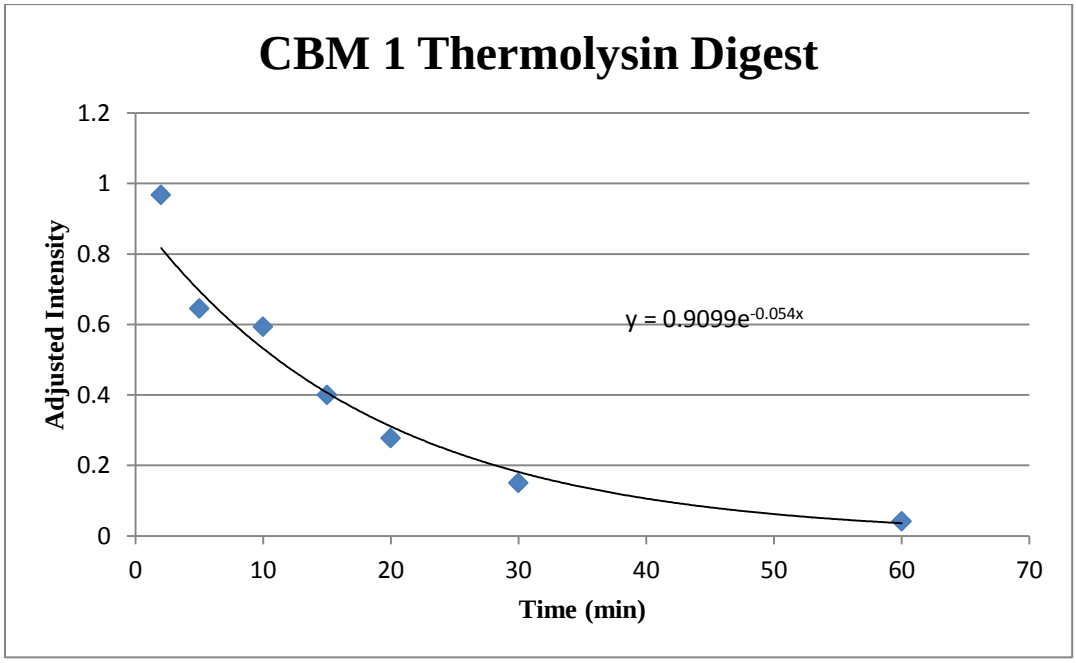

Figure 3.1. The results of the thermolysisn digest of CBM1. The adjusted intensity of each point is the intensity of the sample molecular ion divided by that of the internal standard. The exponential decay fit line and equation that were used to calculate the half-life are displayed.

3.II.b. Results and Discussion

Comparison of the thermolysin half-lives of the nine mono-glycosylated CBMs to the unglycosylated CBM in Figure 3.1A reveals an interesting trend.

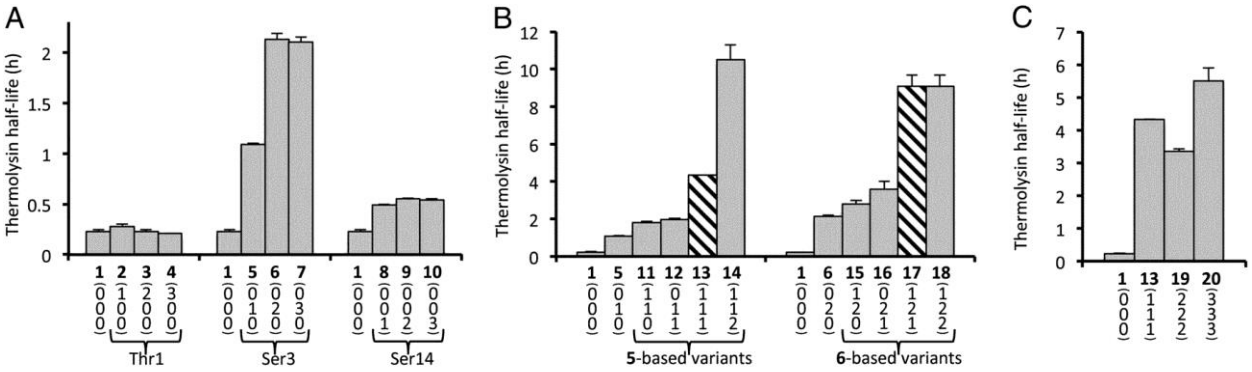

Figure 3.2. The thermolysin half-lives of the 20 CBM glycoforms. A) The nine mono-glycosylated CBM glycoforms divided by glycosylation site. B) The multiply-glycosylated CBM based on additional glycosylation of glycoforms 5 and 6. C) The three CBM glycoforms with all three glycosylation sites occupied by the same glycans. The bars with the hatched pattern represent glycoforms with the best overall (stability and binding) properties.

Glycosylation at Thr-1 did not increase the proteolytic stability of the molecule at all, glycosylation at Ser-14 increased the proteolytic stability a little, and glycosylation at Ser-3 had a large impact on the proteolytic stability. The nonglycosylated CBM 1 has a thermolysin half-life of about 0.2 h, whereas the Ser-3 glycosylated CBM variants, 6 and 7, have half-lives of more than 2 h, an increase of over 10-fold. The greatly increased proteolytic stability from Ser-3 glycosylation is likely due to the site's proximity to a thermolysin cleavage site (see Figure 2.5). Glycosylation so close to the cleavage site may directly shield the protein backbone from

cleavage or sterically hinder protease approach. Thr-1 and Ser-14 are farther away from thermolysin cleavage sites (see Figure 2.5), so it is not surprising that glycosylation at those sites has less of a protective effect. Increasing the size of the glycan on Ser-3 from the monosaccharide to a disaccharide increased the half-life from 1 hour to 2 hours, but going from the disaccharide to the trisaccharide did not increase the half-life. Therefore, glycan size does appear to increase the half-life up until a point.

Figure 3.1 B shows the multiply glycosylated CBM glycoforms formed by adding additional glycans to the Ser-3 glycosylated glycoforms 5 and 6. It is apparent that adding glycans to Thr-1 and Ser-14 after Ser-3 is already glycosylated can greatly increase the proteolytic half-life (up to 11 hours for glycoform 14). This large increase is interesting because Thr-1 and Ser-14 glycosylation on their own have only a small impact on proteolytic stability. It is possible that the multiply glycosylated analogs adopt a fold that is less susceptible to proteolytic degradation.

Interestingly, the most heavily glycosylated CBM glycoform (CBM 20 in Figure 3.1 C), did not have the longest half-life (CBM 20's half-life is only 6 hours compared to the 11 hour half-life of CBM 14). One possible explanation is that too much glycosylation partially undoes the protective effect by disrupting the protein fold, and potentially exposing more or better cleavage sits to the protease. This hypothesis is supported by the fact that the heavily glycosylated CBM 20 has a relatively small amount of β-sheet structure and a relatively high amount of unordered structure compared to the other glycoforms (see Figure 2.7).

The numerical results for the digest studies, with uncertainties, are reported in full in Table app.4.1 in Appendix 4.

3.II.c. Materials and Methods

The thermolysin digestions were performed at 37°C in 100 μL of solution (50 mM Tris-HCl buffer, 0.5 mM $CaCl_2$, pH = 8.0) with an initial CBM concentration of 270 μM. The CBM and thermolysin were initially present in a 20:1 molar ratio. 10 μL aliquots were taken at specific time intervals and quenched with an equal volume of 5% AcOH. Each sample was analyzed by Quantitative MALDI-TOF Mass Spectrometry (described below) to calculate the change in CBM concentration with time. The digestion rate was determined by monitoring and fitting data to the first order exponential decay of the full length CBM glycoform over time.

In the quantative MALDI-TOF experiments, internal reference standard solutions of each CBM glycoform were prepared per experiment by serial dilution (10 μL per concentration). To all sample aliquots, 150 pmol of a CBM internal standard peptide ($\Delta m/z \geq 162$ Da) in H_2O:MeCN:AcOH (1:1:3.3% 3 μL) was added. 0.5 μL of each sample was spotted directly on a 100 well MALDI target plate with 1.126 μL of α-cyano-4-hydroxycinnamic acid (CHCA) matrix (6.2 mg/ml) in MeOH:MeCN:H_2O (36:56:8) and allowed to air dry (~5 min). Spectra were acquired on a Voyager-DE^{TM} STR MALDI-TOF mass spectrometer (Applied Biosystems) in linear positive ion mode, with 50 shots per spectra. The laser intensity was set to 1950, the accelerating voltage was set to 20,000 V, the extraction delay time was 100 ns, and the grid voltage was set to 94%. The low mass gate was set to 500 Da and data were collected from 3200-5000 Da (5500 Da for glycoform **20**). An in-house MATLAB program was written to determine the ratio of analyte ion intensities between the CBM and the CBM internal standard. From these data, a standard linear calibration curve was generated for each experiment to calculate the absolute CBM concentration from CBM:CBM internal standard ion intensity ratios.

3.III. Thermostability

3.III.a. Experimental Design

Variable temperature circular dichroism (CD) was use to investigate whether or not mannosylation affected the thermostability of the CBM glycoforms. CD provides information on a protein's secondary structure by monitoring how the protein affects the ellipticity of circularly polarized light.[52] By monitoring the CD signal from a protein at a specific wavelength as temperature is increased, one can get an idea of how much a protein's structure changes with temperature.[53] If an appropriate temperature range is chosen, the protein will go from its folded to an unfolded state and the data can be fit by a sigmoidal curve. The point of inflection of this sigmoidal curve (where the second derivative of the curve is equal to zero) is the melting point of the protein. At this point, the protein is 50% folded and 50% unfolded. A representative variable temperature CD melt for the unglycosylated CBM is shown in Figure 3.3 below.

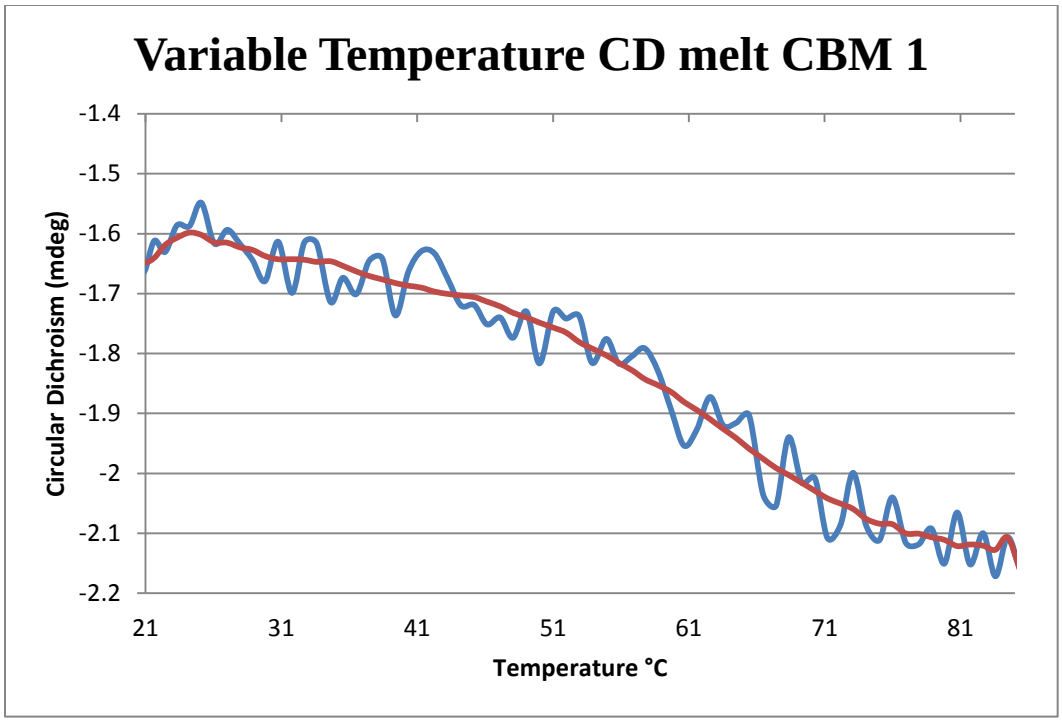

Figure 3.3. The variable temperature CD melt for CBM glycoform 1. Performed between 20 and 90 °C while monitoring at 217 nm in 10 mM NaOAc pH = 5.2. The blue line represents the experimental data and the red line is the sigmoidal fit that was used to calculate the melting temperature.

3.III.b. Results and Discussion

The results show that O-mannosylation affects CBM thermostability in a site-specific manner. Looking at the mono-glycosylated CBM glycoforms on the left of Figure 3.4, we can see that O-mannosylation at Ser-3 leads to the most substantial stabilization, with the increase in melting temperature of 11 °C as compared to the unglycosylated CBM 1. Mannosylation at Ser-14 also leads to noticeable, but less pronounced stabilization than Ser-3 mannosylation. Thr-1 mannosylation displayed the least stabilizing ability, and glycosylation at this site actually slightly destabilized some of the glycoforms. The glycan size does not appear to be directly

related to the magnitude of the stabilizing effect, which is a trend that has been previously noted for protein N-glycosylation.[54]

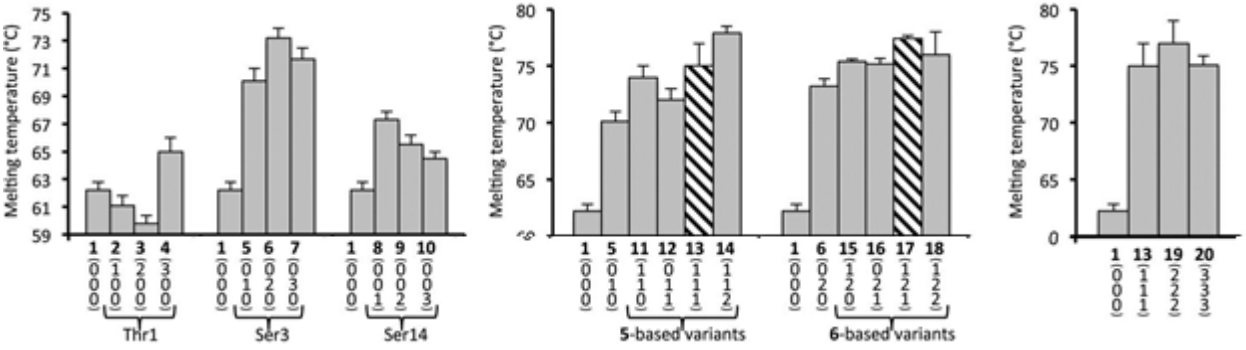

Figure 3.4. The thermal melting point of each CBM glycoform. (Left) The nine mono-glycosylated CBM glycoforms divided by glycosylation site. (Middle) The multiply-glycosylated CBM based on additional glycosylation of glycoforms 5 and 6. (Right) The three CBM glycoforms with all three glycosylation sites occupied by the same glycans. The bars with the hatched pattern represent glycoforms with the best overall (stability and binding) properties.

The correlation between thermostability and glycan density is not very clear (see the middle and left of Figure 3.4). All of the multiply glycosylated CBM glycoforms have relatively high melting pints between 72 and 78°C, but no noticeable trend exists based on the glycosylation patterns. Very high glycan density does not reduce thermostability as it does to an extent with proteolytic stability, but high glycan density does not appear to increase thermostability much when the extent of glycosylation goes beyond a disaccharide at Ser-3.

Mannosylation at Ser-3 leads to the largest increases in thermostability, which suggests that something makes the site different than the other glycosylation sites. A study by Kelly et al.[55] suggested that interactions between Tyr residues at position i and N-linked glycans attached at position i + 2 can lead to substantial thermostabilization in model proteins. This stabilization was purported to be caused by an interaction between the hydrophobic face of the sugar and the aromatic ring of the Tyr residue. Ser-3 in our CBM molecule has a Tyr residue 2 amino acids

away, so it is reasonable to hypothesize that a similar sugar-aromatic ring interaction may be stabilizing our molecule. Further studies are being conducted in the Tan Lab to verify this hypothesis.

The numerical results for the CD melting point studies, with uncertainties, are reported in full in Table app.4.1 in Appendix 4.

3.III.c. Materials and Methods

All variable temperature CD melts were performed using an Applied PhotophysicsChirascan^{TM}-plus CD spectrometer and a 0.5 mm quartz cuvette under nitrogen at a flow rate of 1 L/min. Lyophilized CBM glycoforms were suspended in 10 mM sodium acetate (pH = 5.2) at a concentration of 0.2 mg/ml. The melts were performed by ramping the temperature of the sample from 20 to 94°C at a rate of 1 °C/min while monitoring the CD signal at 217 nm. The melts resulted in roughly sigmoidal melting curves and the point of inflection of the curve was interpreted to be the melting point of the glycoform.[56]

CHAPTER 4

BINDING STUDIES

4.I. Introduction

The next natural course of action was to study the effects of glycosylation on the CBM's natural function. In nature, the CBM binds to cellulose and thus increases the effective concentration of *Tr*Cel7A on the surface of crystalline cellulose. The strength of CBM binding appears to be directly correlated to *Tr*Cel7A activity, with higher affinity CBMs leading to faster and more complete degradation of crystalline cellulose.[57] Previously performed free energy calculations predicted that glycosylation at Thr-1, Ser-3 and Ser-14 would improve the affinity of the *Tr*Cel7A Family 1 CBM for crystalline cellulose,[32] so binding experiments were performed using the CBM glycoforms to see if the predictions were true. Functional studies like these crystalline cellulose binding experiments are essential for glyco-engineering efforts because improving the proteolytic and theromostability of a protein with glycosylation isn't really meaningful if the glycosylation also impairs protein function. Fortunately, glycosylation did not appear to prevent crystalline cellulose binding for most of the glycoforms and some glycoforms with fairly strong binding were identified.

4.II. Experimental Design

An established depletion isotherm method[58] was used to study the binding of the CBM glycoforms to insoluble cellulose crystals known as bacterial microcrystalline cellulose (BMCC).

In short, different concentrations of the CBM analogs were stirred in solution with a fixed amount of BMCC crystals at constant temperature for 2 hours, the solutions were centrifuged to precipitate the cellulose crystals and bound CBM molecules, and the concentration of unbound CBM molecules remaining in each solution was measured. Various methods were used to attempt to quantify the concentration of free CBM in the test solutions after each depletion test, and it turned out that the quantitative MALDI technique used in the digestion study was the best choice given the instrumentation that was available. The technique is quite accurate and it allowed for the use of very small amounts of CBM in each test. Using the free CBM concentrations from each trial, the binding constants were calculated using the equations given in the materials and methods section. Fortunately, Erick Green developed an automated data analysis program that greatly accelerated the binding studies. A representative BMCC binding curve for the unglycosylated CBM 1 is shown in Figure 4.1 below.

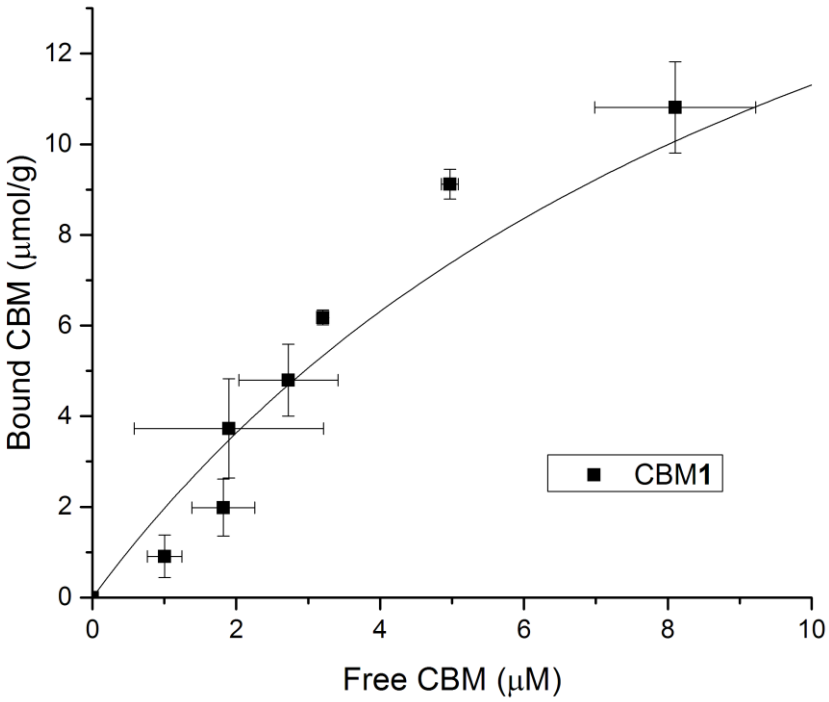

Figure 4.1. The BMCC binding curve of CBM glycoform 1. Data points represent averaged data of at least two trials. Error bars are provided. The fit line was calculated using equation 4.2 in the materials and methods section of this chaper.

The adsorption isoforms for the remaining 19 CBM glycoforms are available in Appendix 5.

4.III. Results and Discussion

The results of the binding study are displayed in Figure 4.2 below. Importantly, the binding affinity observed for the unglycosylated CBM is in agreement with the value observed in other studies,[58–60] which validates the experimental techniques utilized in this study.

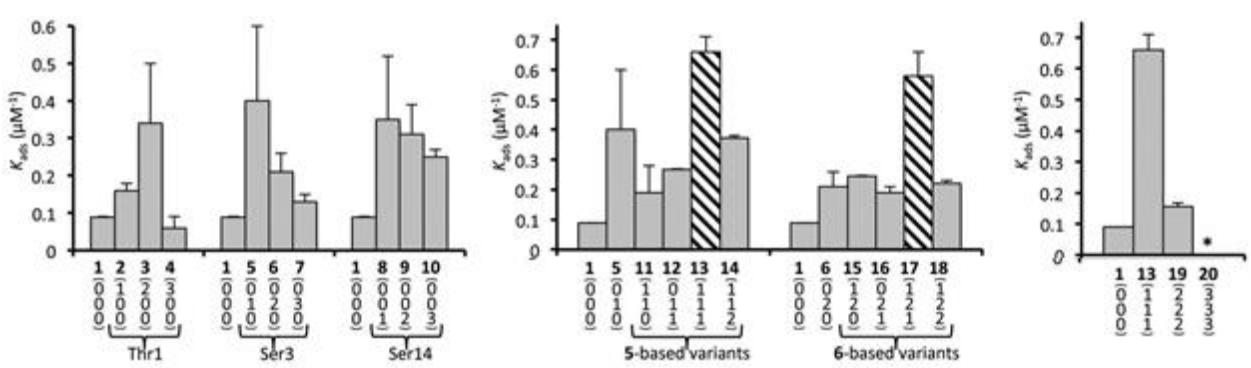

Figure 4.2. The BMCC binding affinity of the CBM glycoforms. (Left) The nine mono-glycosylated CBM glycoforms divided by glycosylation site. (Middle) The multiply-glycosylated CBM based on additional glycosylation of glycoforms 5 and 6. (Right) The three CBM glycoforms with all three glycosylation sites occupied by the same glycans. The bars with the hatched pattern represent glycoforms with the best overall (stability and binding) properties.

Looking at the results for the mono-glycosylated glycoforms, reveals that attachment of a single mannose to Ser-3 and Ser- 14 substantially increases the binding affinity of the CBM for BMCC (by about 4-fold). The addition of mannose di and tri-saccharides to Ser-3 and Ser-14 reduces the BMCC binding affinity relative to the glycoforms with just a monosaccharide at those positions. This result is similar to observations by Boraston *et al*.,[61] who reported that large *N*-linked glycans were detrimental to a Family 2 CBM cellulose binding. These results may be due to steric hindrance. It is possible that small glycans improve binding but longer glycans can interfere with the interactions between the hydrophobic surface of BMCC and the highly conserved Tyr-5, Tyr-31, and Tyr-32 residues on the CBM that are thought to be responsible for binding.[61,62]

Interestingly, the attachment of a single mannose to Thr-1 did not increase the BMCC binding affinity of the CBM much (see 2 in Figure 4.2), but the addition of a mannose-mannose disaccharide increase the BMCC finding affinity by nearly 4-fold just as the mono-saccharides did for Ser-3 and Ser-14. This observation may be due to the greater distance of Thr-1 from the

binding face compared to the other glycosylation sites (see Figure 1.1). Because of the greater distance, a longer glycan is apparently required at Thr-1 in order to increase the binding affinity. The addition of a mannose trisaccharide to Thr-1 decreased the BMCC binding affinity of the CBM, possibly due to a steric hindrance effect as was hypothesized for the longer glycans at Ser-3 and Ser-14. It is also possible that large glycans alter CBM folding in a way that limits binding

The results of the BMCC binding experiments on the multiply glycosylated CBM glycoforms were interesting. As shown in Figure 4.2, the CBM glycoforms with glycans attached to all three sites had higher binding affinities than those with just two attachment sites (compare 13 and 17 to 11, 12, 15 and 16 in Figure 4.2). As with the mono-glycosylated CBM glycoforms, the multiply glycosylated CBM's with smaller glycans had higher binding affinities than those with larger glycans (compare 13 to 19 in Figure 4.2). In fact, glycoform 13, which has a single mannose at Thr-1, Ser-3 and Ser-14, has the strongest binding affinity to BMCC of all of the glycoforms. Apparently, glycosylation does increase BMCC binding affinity as long as the glycans remain small, and it is beneficial to have all three of the glycosylation sites occupied. These results suggest that mannose monosaccharide attachment to the *Tr*Cel7A Family 1 CBM glycosylation sites could be a viable way to engineer enzymes with higher binding affinity and thus improved catalytic function.

The numerical results for the binding studies, with uncertainties, are reported in full in Table app.4.1 in Appendix 4.

4.IV. Materials and Methods

In the BMCC binding experiments, 100-125 μg CBM samples were prepared by lyophilization of the pure synthetic products. Stock CBM solutions were prepared by diluting sample to 0.5 mg/ml concentration with acetate buffer (50 mM NaOAc, 50 mM NaCl, pH 5.0). The stock solutions were serially diluted to produce 1, 2, 3, 4, 8, 12, 16, 20, 30, 40 μM CBM samples which were two fold diluted with either acetate buffer (to produce standard curves) or bacterial crystalline cellulose obtained from *Acetobacter xylinus sub sp. Sucrofermentans* suspended in acetate buffer (to obtain experimental results). The samples were stirred in 2 mL Eppendorf tubes with 5mm magnetic stir bars for 2 hours at 4°C and 1100 rpm. The samples were then centrifuged at 14,000 *x g* for 10 minutes at 4°C to remove the BMCC and bound cellulose from solution. From each sample two 10 μl aliquots were taken (from the top of the solution) and mixed with 3 μL (~150pmol) of a MALDI-MS standard CBM in MeCN:H_2O:AcOH (1:1:5%). Samples were stored at 4°C until MALDI-TOF data acquisition.

1.1 μl of Alpha-Cyano-4-hydroxycinnamic acid (6.2 mg/ml in 36% MeOH, 56% MeCN, and 8% H_2O) was spotted on a gold 100 well MALDI target plate followed immediately by 0.5 μl of each binding sample. The spots were allowed to dry for a minimum of 5 minutes. MALDI-TOF-MS data was acquired on a Voyager-DE™ STR mass spectrometer (Applied Biosystems) in linear mode, with 50 shots/spectrum, a mass range of 3200-5000 Da (3200-5500 Da for CBM 20), and a laser intensity of 1950. For the standard samples, 2-3 spectra were acquired per spot and averaged. For the experimental samples, 1-2 spectra were acquired per spot and averaged.

All spectra were exported as text files for data analysis. An in house MATLAB code developed by Erick Greene was used to determine the free concentration of CBM in the binding solutions by comparing the intensities of the experimental and standard mass peaks and then comparing the result to a standard curve (obtained using known CBM concentrations). The

concentration of free CBM was then used to determine the amount of bound CBM using Equation 4.1.

Equation 4.1: $$Bound\ \left(\frac{\mu mol}{g}\right) = \frac{[Initial\ (\mu M) - Free\ (\mu M)] \times Vol\ (L)}{Cellulose\ (g)}$$

Where bound represents the concentration of CBM bound to crystalline cellulose per unit cellulose, initial represents in the initial concentration of CBM present in the binding mixture, free represents the concentration of unbound CBM at equilibrium, and vol represents the total volume of the binding solution (100 μL for these experiments).

The resulting data were plotted in Origin 9 Pro and nonlinearly curve fitted to the single-site Langmuir Equation (Equation 4.2) to determine binding constants.

Equation 4.2: $$[Bound] = \frac{N_o \times [Free]}{K_{ads}^{-1} + [Free]}$$

Where N_o represents the binding capacity of the CBM glycoform to crystalline cellulose and K_{ads} is the association constant of the CBM-cellulose complex.

CHAPTER 5

CONCLUSIONS

5.I. The Effects of Glycosylation on *Tr*Cel7A's Family 1 CBM

The synthetic glycoform library conclusively demonstrated that *O*-glycosylation enhances the stability and cellulose binding affinity of the *Tr*Cel7A Family 1 CBM. This study further demonstrates how chemical synthesis can be used as a tool to study the effects of glycosylation. The results clearly show that Ser-3 is the glycosylation site with the largest impact on CBM stability. Considering previous studies on *N*-glycosylation,[49,63] it is reasonable to hypothesize that the large enhancements caused by *O*-mannosylation at Ser3 might be at least partially due to interactions between the mannose residues and nearby Tyr-5. Further studies are being conducted in the Tan Lab to explore this hypothesis. Analysis of the data also revealed two glycoforms—13 and 17, shown with a hatched pattern in Figure 3.2, 3.4 and 4.2—with greatly enhanced proteolytic stability, theromostability and BMCC binding affinity. These desirable glycoforms, which have small glycans at all three glycosylation sites, may potentially find use in biofuel production applications if an appropriate expression system can be identified.

Furthermore, the results demonstrate that excessive glycosylation can be detrimental to CBM stability and function. This statement is certainly true for proteolytic stability where the most heavily glycosylated glycoform (CBM 20) had a half live of only 5.5 hours whereas glycoforms with more intermediate amounts of glycosylation such as glycoforms 14, 17 and 18 had half-lives between 9 and 10.5 hours. It is possible that too much glycosylation has less of a protective effect because it partially disrupts the protein fold which may expose more or better

cleavage sits to the protease. A similar trend can be observed for BMCC binding affinity where glycoform 13, which has a monosaccharide at Thr-1, Ser-3 and Ser-14, has the highest binding affinity of all the analogs. Glycoform 19 has a disaccharide at all of the positions, and it has an affinity of less than 1/3 that of glycoform 13. This result again suggests that a moderate amount of glycosylation is most beneficial. In the case of binding affinity, the reduced binding of heavily glycosylated glycoforms is likely due to sterics. The large glycans probably interfere with the binding interaction between the three Tyr residues on the CBM binding face and the crystalline surface of BMCC.[61,62] Interestingly, very high glycan density does not reduce thermostability. However, high glycan density does not appear to increase thermostability much when the extent of glycosylation goes beyond a disaccharide at Ser-3. Apparently excessive glycosylation does not limit thermostability as it does the other two properties that were studied. Therefore, it is necessary to test several different glyforms in order to determine the glycosylation pattern that provides the best balance between stability and function. Such studies allow for the identification of glycoforms with the best overall properties such as glycoforms 13 and 17 in this study (shown with a hatched pattern in Figure 3.2, 3.4 And 4.2).

5.II. CBM Glycosylation and Biofuels Production

The results of this study demonstrate that CBM mannosylation confers several properties that are potentially beneficial for CBM applications in biofuel production. For instance, the enhanced proteolytic stability of the mannosylated CBM glycoforms could help protect them from the proteases that fungi secrete during biomass depolymerization.[41] Such protection would reduce the amount of CBM and cellulase lost during the digestion of cellulose and thus increase

biomass conversion rates and yields. Thermostability is also a desirable trait of industrial enzymes as illustrated by the fact that many studies have engineered cellulases for improved thermostability through amino acid substitutions or through domain and sequence shuffling.[45,64] Thermostability is beneficial because it allows biofermentations to be run at higher temperatures with less enzyme loss. Our results indicate that CBM O-mannosylation is an effective way to increase CBM theromostability by up to 16 °C.

The BMCC binding affinity results are especially promising in terms of biofuels applications. As previously mentioned, the strength of CBM binding appears to be directly correlated to *Tr*Cel7A activity, with higher affinity CBMs leading to faster and more complete degradation of crystalline cellulose.[57] The results show that glycosylated CBMs have binding affinities up to 7.4 times greater than the unglycosylated protein, which suggest that O-mannosylation is a viable way to increase BMCC binding affinity and thus the catalytic efficiency of cellulose depolymerization by *Tr*Cel7A. Indeed, the affinity enhancements observed in this study are similar to those reported by other researchers who attempted to increase CBM binding affinity through amino acid mutations.[21,57] The added benefit of glycosylation over such mutations is that it also improves the stability properties of the protein.

It is important to note that the results of this study have implications that go beyond *Tr*Cel7A. The *O*-linked glycosylation sites examined in this study are highly conserved across Family 1 CBMs,[31] suggesting that the enhanced properties from glycosylation observed here likely occur throughout this ubiquitous CBM family. Therefore, glycosylation may be a means to improve the properties of a wide range of cellulase CBM's, and it may find several applications in biofuel production.

This study illustrates that chemical synthesis is a viable way to produce glycoform libraries for the systematic study of glycosylation. Here, there result was the identification of two glycoforms with excellent properties for biofuel production applications. Additional work will have to be done in order to develop expression systems that favor the production of the desired glycoforms. In the future, it should be possible to use this approach to develop glycoengineering strategies for other industrially or therapeutically important proteins.

REFERENCES

(1) Brooks, S. a Strategies for analysis of the glycosylation of proteins: current status and future perspectives. *Molecular biotechnology* **2009**, *43*, 76–88.

(2) Shental-Bechor, D.; Levy, Y. Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Current opinion in structural biology* **2009**, *19*, 524–33.

(3) Grünewald, S. Congenital disorders of glycosylation: rapidly enlarging group of (neuro)metabolic disorders. *Early human development* **2007**, *83*, 825–30.

(4) Anthony, R. M.; Nimmerjahn, F.; Ashline, D. J.; Reinhold, V. N.; Paulson, J. C.; Ravetch, J. V Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science (New York, N.Y.)* **2008**, *320*, 373–6.

(5) Seipert, R. R.; Dodds, E. D.; Lebrilla, C. B. Exploiting Differential Dissociation Chemistries of O-Linked Glycopeptide Ions for the Localization of Mucin-Type Protein Glycosylation research articles. **2009**, 493–501.

(6) Grogan, M. J.; Pratt, M. R.; Marcaurelle, L. a; Bertozzi, C. R. Homogeneous glycopeptides and glycoproteins for biological investigation. *Annual review of biochemistry* **2002**, *71*, 593–634.

(7) Nilsson, B. L.; Soellner, M. B.; Raines, R. T. Chemical synthesis of proteins. *Annual review of biophysics and biomolecular structure* **2005**, *34*, 91–118.

(8) Yuan, Y.; Chen, J.; Wan, Q.; Wilson, R. M.; Danishefsky, S. J. Toward fully synthetic, homogeneous glycoproteins: advances in chemical ligation. *Biopolymers* **2010**, *94*, 373–84.

(9) Sakamoto, I.; Tezuka, K.; Fukae, K.; Ishii, K.; Taduru, K.; Maeda, M.; Ouchi, M.; Yoshida, K.; Nambu, Y.; Igarashi, J.; Hayashi, N.; Tsuji, T.; Kajihara, Y. Chemical synthesis of homogeneous human glycosyl-interferon-β that exhibits potent antitumor activity in vivo. *Journal of the American Chemical Society* **2012**, *134*, 5428–31.

(10) Wang, P.; Dong, S.; Brailsford, J. a; Iyer, K.; Townsend, S. D.; Zhang, Q.; Hendrickson, R. C.; Shieh, J.; Moore, M. a S.; Danishefsky, S. J. At last: erythropoietin as a single glycoform. *Angewandte Chemie (International ed. in English)* **2012**, *51*, 11576–84.

(11) Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science (New York, N.Y.)* **2007**, *315*, 804–7.

(12) Wilson, D. B. Cellulases and biofuels. *Current opinion in biotechnology* **2009**, *20*, 295–9.

(13) Lynd, L. R.; Weimer, P. J.; Zyl, W. H. Van; Isak, S.; Pretorius, I. S. Microbial Cellulose Utilization : Fundamentals and Biotechnology Microbial Cellulose Utilization : Fundamentals and Biotechnology. **2002**, *66*.

(14) Bayer, E. a; Belaich, J.-P.; Shoham, Y.; Lamed, R. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annual review of microbiology* **2004**, *58*, 521–54.

(15) Demain, A. L.; Newcomb, M.; Wu, J. H. D. Cellulase, Clostridia, and Ethanol†. **2005**, *69*, 124–154.

(16) Fontes, C. M. G. a; Gilbert, H. J. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annual review of biochemistry* **2010**, *79*, 655–81.

(17) Brunecky, R.; Alahuhta, M.; Xu, Q.; Donohoe, B. S.; Crowley, M. F.; Kataeva, I. a; Yang, S.-J.; Resch, M. G.; Adams, M. W. W.; Lunin, V. V; Himmel, M. E.; Bomble, Y. J. Revealing nature's cellulase diversity: the digestion mechanism of Caldicellulosiruptor bescii CelA. *Science (New York, N.Y.)* **2013**, *342*, 1513–6.

(18) Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical journal* **2004**, *382*, 769–81.

(19) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **2014**, *42*, D490–5.

(20) Kraulis, P. J.; Clare, G. M.; Nilges, M.; Jones, T. A.; Pettersson, G.; Knowles, J.; Gronenborn, A. M. Domain of Cellobiohydrolase I from Trichoderma reesei . A Study Using Nuclear Annealingt. **1989**, 7241–7257.

(21) Linder, M.; Lindeberg, G.; Reinikainen, T.; Teeri, T. T.; Pettersson, G. The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. *FEBS letters* **1995**, *372*, 96–8.

(22) Bu, L.; Nimlos, M. R.; Shirts, M. R.; Ståhlberg, J.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. Product binding varies dramatically between processive and nonprocessive cellulase enzymes. *The Journal of biological chemistry* **2012**, *287*, 24807–13.

(23) Lehtiö, J.; Sugiyama, J.; Gustavsson, M.; Fransson, L.; Linder, M.; Teeri, T. T. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding

modules. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 484–9.

(24) Deshpande, N.; Wilkins, M. R.; Packer, N.; Nevalainen, H. Protein glycosylation pathways in filamentous fungi. *Glycobiology* **2008**, *18*, 626–37.

(25) Beckham, G. T.; Dai, Z.; Matthews, J. F.; Momany, M.; Payne, C. M.; Adney, W. S.; Baker, S. E.; Himmel, M. E. Harnessing glycosylation to improve cellulase activity. *Current opinion in biotechnology* **2012**, *23*, 338–45.

(26) Hui, J. P. M.; Lanthier, P.; White, T. C.; McHugh, S. G.; Yaguchi, M.; Roy, R.; Thibault, P. Characterization of cellobiohydrolase I (Cel7A) glycoforms from extracts of Trichoderma reesei using capillary isoelectric focusing and electrospray mass spectrometry. *Journal of Chromatography B: Biomedical Sciences and Applications* **2001**, *752*, 349–368.

(27) Stals, I.; Sandra, K.; Geysens, S.; Contreras, R.; Van Beeumen, J.; Claeyssens, M. Factors influencing glycosylation of Trichoderma reesei cellulases. I: Postsecretorial changes of the O- and N-glycosylation pattern of Cel7A. *Glycobiology* **2004**, *14*, 713–24.

(28) Pinto, R.; Carvalho, J.; Mota, M.; Gama, M. Large-scale production of cellulose-binding domains. Adsorption studies using CBD-FITC conjugates. *Cellulose* **2006**, *13*, 557–569.

(29) Payne, C. M.; Resch, M. G.; Chen, L.; Crowley, M. F.; Himmel, M. E.; Taylor, L. E.; Sandgren, M.; Ståhlberg, J.; Stals, I.; Tan, Z.; Beckham, G. T. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 14646–51.

(30) Harrison, M. J.; Nouwens, a S.; Jardine, D. R.; Zachara, N. E.; Gooley, a a; Nevalainen, H.; Packer, N. H. Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of Trichoderma reesei. *European journal of biochemistry / FEBS* **1998**, *256*, 119–27.

(31) Beckham, G. T.; Matthews, J. F.; Bomble, Y. J.; Bu, L.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R.; Crowley, M. F. Identification of amino acids responsible for processivity in a Family 1 carbohydrate-binding module from a fungal cellulase. *The journal of physical chemistry. B* **2010**, *114*, 1447–53.

(32) Taylor, C. B.; Talib, M. F.; McCabe, C.; Bu, L.; Adney, W. S.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. Computational investigation of glycosylation effects on a family 1 carbohydrate-binding module. *The Journal of biological chemistry* **2012**, *287*, 3147–55.

(33) Chen, L.; Tan, Z. A convenient and efficient synthetic approach to mono-, di-, and tri-O-mannosylated Fmoc amino acids. *Tetrahedron Letters* **2013**, *54*, 2190–2193.

(34) Chen, L.; Drake, M. R.; Resch, M. G.; Greene, E. R.; Himmel, M. E.; Chaffey, P. K.; Beckham, G. T.; Tan, Z. Specificity of O-glycosylation in enhancing the stability and cellulose binding affinity of Family 1 carbohydrate-binding modules. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, 1–6.

(35) Fields, GB Nobel, R. Solid phase peptide synthesis utilizing 9-fluprenylmethoxycarbonyl amino acids. *Int. J. Pept. Protein Res.* **1990**, *35*, 161–214.

(36) Tan, Z.; Shang, S.; Halkina, T.; Yuan, Y.; Danishefsky, S. J. Toward Homogeneous Erythropoietin : Non-NCL-Based Chemical Synthesis of the Gln 78 -Arg 166 Glycopeptide Domain. **2009**, 5424–5431.

(37) Johansson, G Stahlberg, Lindenberg,G Engstrom, A Pettersson, G. Isolated fungal cellulase terminal domains and a synthetic minimum analogue bind to cellulose. **1989**, *243*, 389–393.

(38) Mattinen, M. L.; Kontteli, M.; Kerovuo, J.; Linder, M.; Annila, a; Lindeberg, G.; Reinikainen, T.; Drakenberg, T. Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from Trichoderma reesei. *Protein science : a publication of the Protein Society* **1997**, *6*, 294–303.

(39) Hagspiel, K.; Haab, D.; Kubicek, C. P. Applied ° . d Microbiology Biotechnology Protease activity and proteolytic modification of cellulases from a Trichoderma reesei Q M 9414 selectant. **1989**, 61–67.

(40) Haab, D.; Hagspiel, K.; Szakmary, K.; Kubicek, C. P. Formation of the extracellular proteases from Trichoderma reesei QM 9414 involved in cellulase degradation. *Journal of Biotechnology* **1990**, *16*, 187–198.

(41) Schuster, A.; Schmoll, M. Biology and biotechnology of Trichoderma. *Applied microbiology and biotechnology* **2010**, *87*, 787–99.

(42) Dienes, D.; Börjesson, J.; Hägglund, P.; Tjerneld, F.; Lidén, G.; Réczey, K.; Stålbrand, H. Identification of a trypsin-like serine protease from Trichoderma reesei QM9414. *Enzyme and Microbial Technology* **2007**, *40*, 1087–1094.

(43) Banerjee, G.; Scott-Craig, J. S.; Walton, J. D. Improving Enzymes for Biomass Conversion: A Basic Research Perspective. *BioEnergy Research* **2010**, *3*, 82–92.

(44) Yike, I. Fungal proteases and their pathophysiological effects. *Mycopathologia* **2011**, *171*, 299–323.

(45) Dana, C. M.; Saija, P.; Kal, S. M.; Bryan, M. B.; Blanch, H. W.; Clark, D. S. Biased clique shuffling reveals stabilizing mutations in cellulase Cel7A. *Biotechnology and bioengineering* **2012**, *109*, 2710–9.

(46) Viikari, L.; Terhi, A. Thermostable Enzymes in Lignocellulose Hydrolysis. **2007**, 121–145.

(47) Kim, T.-W.; Chokhawala, H. a; Nadler, D. C.; Nadler, D.; Blanch, H. W.; Clark, D. S. Binding modules alter the activity of chimeric cellulases: Effects of biomass pretreatment and enzyme source. *Biotechnology and bioengineering* **2010**, *107*, 601–11.

(48) Hall, M.; Rubin, J.; Behrens, S. H.; Bommarius, A. S. The cellulose-binding domain of cellobiohydrolase Cel7A from Trichoderma reesei is also a thermostabilizing domain. *Journal of biotechnology* **2011**, *155*, 370–6.

(49) Price, J. L.; Powers, D. L.; Powers, E. T.; Kelly, J. W. Glycosylation of the enhanced aromatic sequon is similarly stabilizing in three distinct reverse turn contexts. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 14127–32.

(50) Boyer, A. E.; Gallegos-Candela, M.; Lins, R. C.; Kuklenyik, Z.; Woolfitt, A.; Moura, H.; Kalb, S.; Quinn, C. P.; Barr, J. R. Quantitative mass spectrometry for bacterial protein toxins--a sensitive, specific, high-throughput tool for detection and diagnosis. *Molecules (Basel, Switzerland)* **2011**, *16*, 2391–413.

(51) Wu, J.; Chatman, K.; Harris, K.; Siuzdak, G. An automated MALDI mass spectrometry approach for optimizing cyclosporin extraction and quantitation. *Analytical chemistry* **1997**, *69*, 3767–71.

(52) Greenfield, N. J. Using circular dichrosim spectra to estimate protein secondary structure. *Nat Protoc.* **2006**, *1*, 2876–2890.

(53) Price, J. L.; Shental-bechor, D.; Dhar, A.; Turner, M. J.; Evan, T.; Gruebele, M.; Levy, Y.; Kelly, J. W. Context-Dependent Effects of Asparagine Glycosylation on Pin WW Folding Kinetics and Theromodynamics. *J Am Chem Soc.* **2010**, *132*, 15359–15367.

(54) Hanson, S. R.; Culyba, E. K.; Hsu, T.-L.; Wong, C.-H.; Kelly, J. W.; Powers, E. T. The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 3131–6.

(55) Price, J. L.; Culyba, E. K.; Chen, W.; Murray, A. N.; Hanson, S. R.; Wong, C.-H.; Powers, E. T.; Kelly, J. W. N-glycosylation of enhanced aromatic sequons to increase glycoprotein stability. *Biopolymers* **2012**, *98*, 195–211.

(56) Voutilainen, S. P.; Nurmi-Rantala, S.; Penttilä, M.; Koivula, A. Engineering chimeric thermostable GH7 cellobiohydrolases in Saccharomyces cerevisiae. *Applied microbiology and biotechnology* **2014**, *98*, 2991–3001.

(57) Takashima, S.; Ohno, M.; Hidaka, M.; Nakamura, A.; Masaki, H.; Uozumi, T. Correlation between cellulose binding and activity of cellulose-binding domain mutants of Humicola grisea cellobiohydrolase 1. *FEBS letters* **2007**, *581*, 5891–6.

(58) Linder, M.; Mattinen, M. L.; Kontteli, M.; Lindeberg, G.; Ståhlberg, J.; Drakenberg, T.; Reinikainen, T.; Pettersson, G.; Annila, a Identification of functionally important amino acids in the cellulose-binding domain of Trichoderma reesei cellobiohydrolase I. *Protein science : a publication of the Protein Society* **1995**, *4*, 1056–64.

(59) Taylor, C. B.; Payne, C. M.; Himmel, M. E.; Crowley, M. F.; McCabe, C.; Beckham, G. T. Binding site dynamics and aromatic-carbohydrate interactions in processive and non-processive family 7 glycoside hydrolases. *The journal of physical chemistry. B* **2013**, *117*, 4924–33.

(60) Guo, J.; Catchmark, J. M. Binding specificity and thermodynamics of cellulose-binding modules from Trichoderma reesei Cel7A and Cel6A. *Biomacromolecules* **2013**, *14*, 1268–77.

(61) Boraston, A. B.; Warren, R. A. J.; Kilburn, D. G. carbohydrate-binding module from Cellulomonas fimi : a functional and mutational analysis. **2001**, *430*, 423–430.

(62) Boraston, A. B.; Sandercock, L. E.; Warren, R. A. J.; Kilburn, D. G. O-Glycosylation of a Recombinant Carbohydrate-Binding Module Mutant Secreted by *Pichia pastoris*. *Journal of Molecular Microbiology and Biotechnology* **2003**, *5*, 29–36.

(63) Ashida, H.; Ozawa, H.; Fujita, K.; Suzuki, S.; Yamamoto, K. Syntheses of mucin-type O-glycopeptides and oligosaccharides using transglycosylation and reverse-hydrolysis activities of Bifidobacterium endo-alpha-N-acetylgalactosaminidase. *Glycoconjugate journal* **2010**, *27*, 125–32.

(64) Heinzelman, P.; Snow, C. D.; Wu, I.; Nguyen, C.; Villalobos, A.; Govindarajan, S.; Minshull, J.; Arnold, F. H. A family of thermostable fungal cellulases created by structure-guided recombination. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 5610–5.

APPENDICES

App.1. Synthetic Details and Mass Data for CBM Library

CBM glycoform 1: The unglycosylated peptide was synthesized on a 0.05 mmol scale. After SPPS, 168.2 mg of the crude peptide was obtained. 16 mg (4.28 μmol) of the crude peptide was dissolved in 80 ml of folding buffer and stirred at room temperature for 12 h. After concentration and HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min, 5.18 mg of **1** was obtained as a white solid (30% yield based on resin loading).

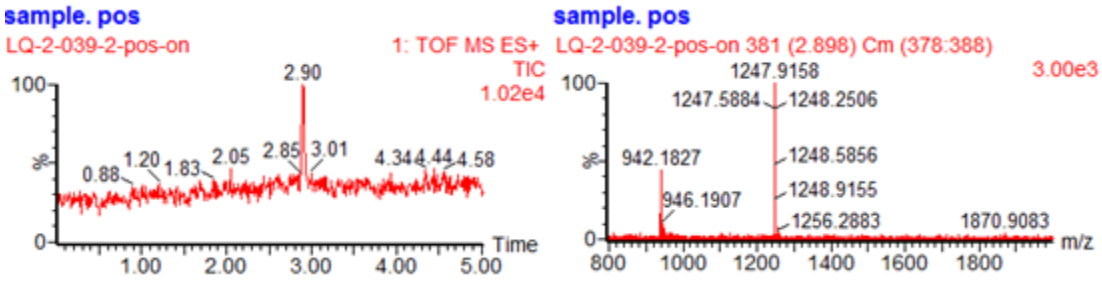

Figure App.1.1. LC-MS trace and ESI-MS of CBM glycoform 1. MS (ESI) Calcd for $C_{159}H_{235}N_{43}O_{54}S_4$: $[M+2H]^{2+}$ m/z = 1870.29, $[M+3H]^{3+}$ m/z = 1247.19, $[M+4H]^{4+}$ m/z = 935.65.

CBM glycoform 2: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 188.7 mg of the crude glycopeptide was obtained. 7 mg (1.71 μmol) of it was dissolved in 450 μl of the solution of hydrazine and stirred at room temperature for 30 min. The reaction was quenched with the solution of AcOH and the pH was adjusted to ~8. The resulting solution was diluted by the addition of 40 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in

H_2O over 30 min afforded the correctly folded **2** (1.04 mg, white solid, 15% yield based on resin loading).

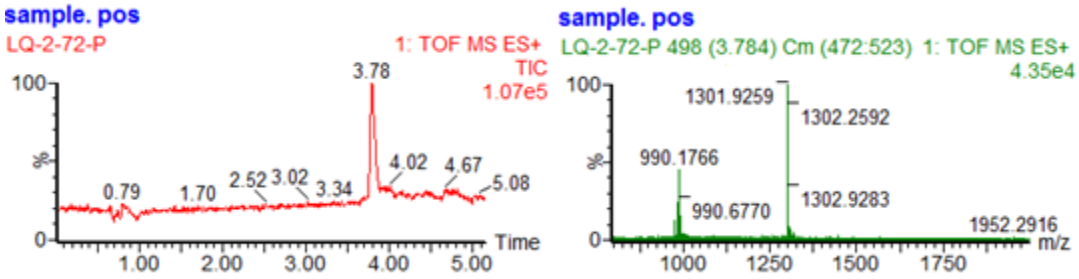

Figure App.1.2. LC-MS trace and ESI-MS of CBM glycoform 2. MS (ESI) Calcd for $C_{165}H_{245}N_{43}O_{59}S_4$: $[M+2H]^{2+}$ m/z = 1951.32, $[M+3H]^{3+}$ m/z = 1301.21, $[M+4H]^{4+}$ m/z = 976.16.

CBM glycoform 3: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 164.4 mg of the crude product was obtained. 16 mg (3.67 µmol) of it was dissolved in 1000 µL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with acetic acid and the pH was adjusted to between 8 and 9. The resulting solution was diluted by the addition of 80 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **3** (0.64 mg, white solid, 3% yield based on resin loading).

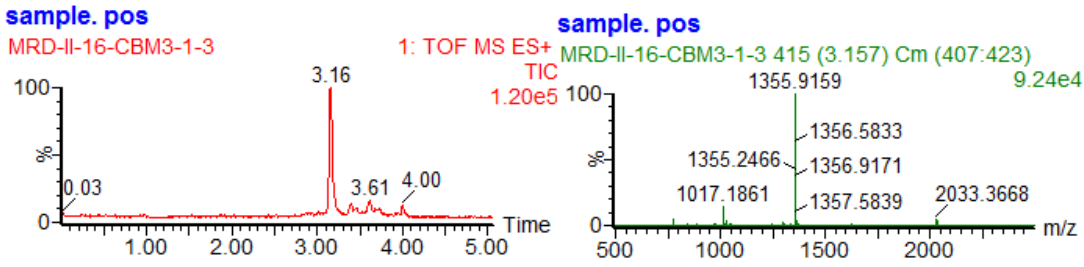

Figure App.1.3. LC-MS trace and ESI-MS of CBM glycoform 3. MS (ESI) Calcd for $C_{171}H_{255}N43O_{64}S_4$: $[M+2H]^{2+}$ m/z = 2032.35, $[M+3H]^{3+}$ m/z = 1355.23, $[M+4H]^{4+}$ m/z = 1016.67.

CBM glycoform 4: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 201.6 mg of the crude product was obtained. 16 mg (3.44 μmol) of it was dissolved in 1000 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the AcOH solution and the pH was adjusted to between 8 and 9. The resulting solution was diluted by the addition of 80 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **4** (1.95 mg, white solid, 12% yield based on resin loading).

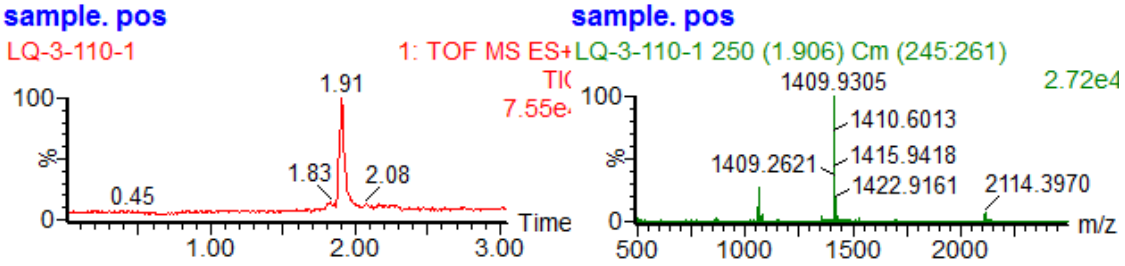

Figure App.1.4. LC-MS trace and ESI-MS of CBM glycoform 4. MS (ESI) Calcd for $C_{177}H_{265}N_{43}O_{69}S_4$: $[M+2H]^{2+}$ m/z = 2113.37, $[M+3H]^{3+}$ m/z = 1409.25, $[M+4H]^{4+}$ m/z = 1057.19.

CBM glycoform 5: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 179.8 mg of the crude glycopeptide was obtained. 16 mg (3.90 μmol) of it was dissolved in 1000 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. Then resulting solution was diluted by the addition of 16 mL folding buffer. After folding and centrifugal concentration, HPLC purification with a linear

gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **5** (1.12 mg, white solid, 6% yield based on resin loading).

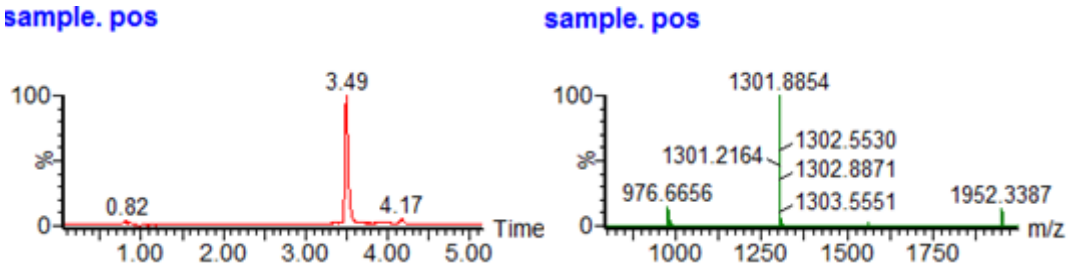

Figure App.1.5. LC-MS trace and ESI-MS of CBM glycoform 5. MS (ESI) Calcd for $C_{165}H_{245}N_{43}O_{59}S_4$: $[M+2H]^{2+}$ m/z = 1951.32, $[M+3H]^{3+}$ m/z = 1301.21, $[M+4H]^{4+}$ m/z = 976.16.

CBM glycoform 6: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 195.0 mg of the crude glycopeptide was obtained. 16 mg (3.67 μmol) of it was dissolved in 1000 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was diluted by the addition of 16 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **6** (1.00 mg, white solid, 6% yield based on resin loading).

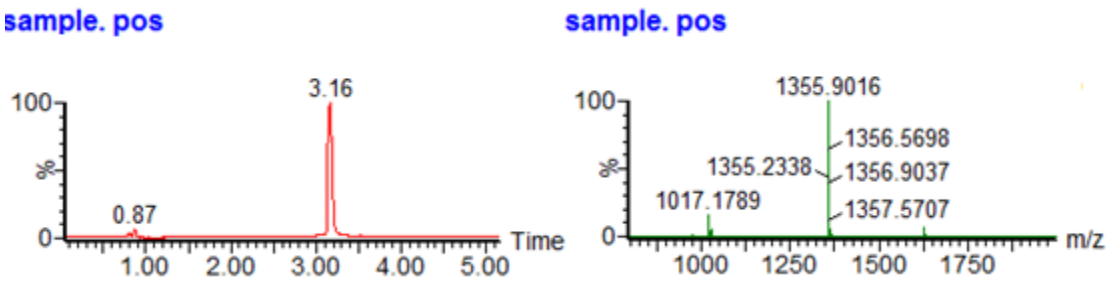

Figure App.1.6. LC-MS trace and ESI-MS of CBM glycoform 6. MS (ESI) Calcd for $C_{171}H_{255}N_{43}O_{64}S_4$: $[M+2H]^{2+}$ m/z = 2032.35, $[M+3H]^{3+}$ m/z = 1355.23, $[M+4H]^{4+}$ m/z = 1016.67.

CBM glycoform 7: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 174.0 mg of the crude product was obtained. 16 mg (3.44 μmol) of the crude material was dissolved in 1000 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the AcOH solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was diluted by the addition of 16 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **7** (1.09 mg, white solid, 6% yield based on resin loading).

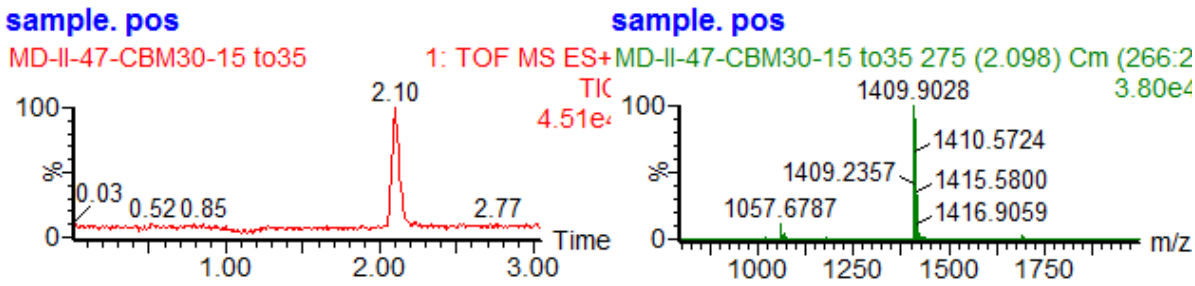

Figure App.1.7. LC-MS trace and ESI-MS of CBM glycoform 7. MS (ESI) Calcd for $C_{177}H_{265}N_{43}O_{69}S_4$: $[M+2H]^{2+}$ m/z = 2113.37, $[M+3H]^{3+}$ m/z = 1409.25, $[M+4H]^{4+}$ m/z = 1057.19.

CBM glycoform 8: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 179.3mg of the crude glycopeptide was obtained. 16 mg (3.90 μmol) of it was stirred in 1000 μL of the hydrazine solution for 30min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. Then resulting glycopeptide solution was diluted by the addition of 16 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **8** (1.25 mg, white solid, 7% yield based on resin loading).

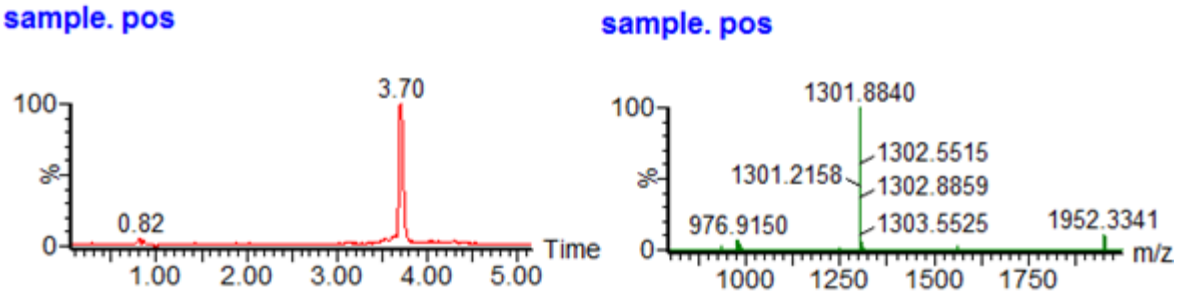

Figure App.1.8. LC-MS trace and ESI-MS of CBM glycoform 8. MS (ESI) Calcd for $C_{165}H_{245}N_{43}O_{59}S_4$: $[M+2H]^{2+}$ m/z = 1951.32, $[M+3H]^{3+}$ m/z = 1301.21, $[M+4H]^{4+}$ m/z = 976.16.

CBM glycoform 9: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 134.7 mg of the crude product was obtained. 16 mg (3.67 μmol) of it was stirred in 1000 μL of the hydrazine hydrate solution for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was diluted by 16 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in

H_2O over 30 min afforded the correctly folded **9** (0.98 mg, white solid, 4% yield based on resin loading).

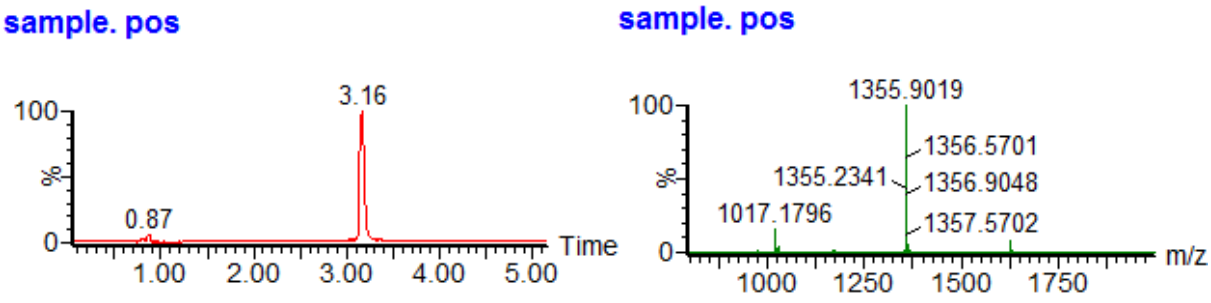

Figure App.1.9. LC-MS trace and ESI-MS of CBM glycoform 9. MS (ESI) Calcd for $C_{171}H_{255}N_{43}O_{64}S_4$: $[M+2H]^{2+}$ m/z = 2032.35, $[M+3H]^{3+}$ m/z = 1355.23, $[M+4H]^{4+}$ m/z = 1016.67.

CBM glycoform 10: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 202.0 mg of the crude glycopeptide was obtained. 16 mg (3.44 μmol) of the crude material was dissolved in 1000 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the AcOH solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 16 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **10** (2.88 mg, white solid, 17% yield based on resin loading).

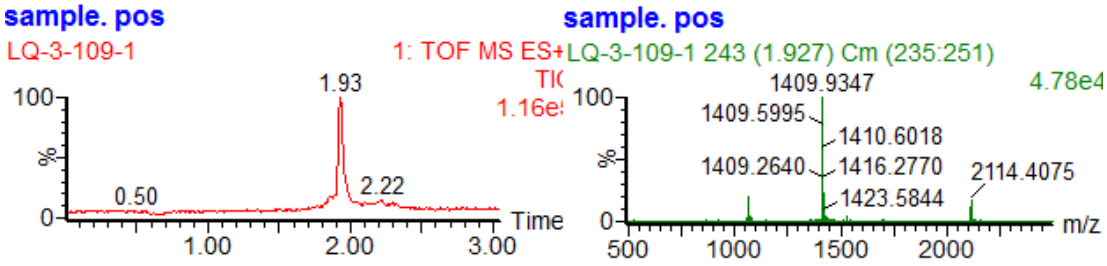

Figure App.1.10. LC-MS trace and ESI-MS of CBM glycoform 10. MS (ESI) Calcd for $C_{177}H_{265}N_{43}O_{69}S_4$: $[M+2H]^{2+}$ m/z = 2113.37, $[M+3H]^{3+}$ m/z = 1409.25, $[M+4H]^{4+}$ m/z = 1057.19.

CBM glycoform 11: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 225.8 mg of the crude glycopeptide was obtained. 3.5 mg (0.795 μmol) of the crude material was stirred in 200 μL of the hydrazine solution at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting solution was then diluted by 16 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **11** (0.45 mg, white solid, 14% yield based on resin loading).

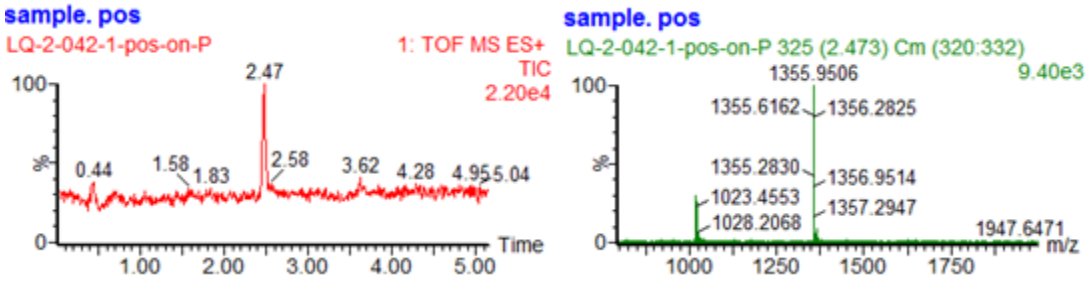

Figure App.1.11. LC-MS trace and ESI-MS of CBM glycoform 11. MS (ESI) Calcd for $C_{171}H_{255}N_{43}O_{64}S_4$: $[M+2H]^{2+}$ m/z = 2032.35, $[M+3H]^{3+}$ m/z = 1355.23, $[M+4H]^{4+}$ m/z = 1016.67.

CBM glycoform 12: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 250 mg of the crude glycopeptide was obtained. 8.0 mg (1.817 μmol) of the crude material was dissolved in 450 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 40 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **12 (**1.33 mg, white solid, 20% yield based on resin loading).

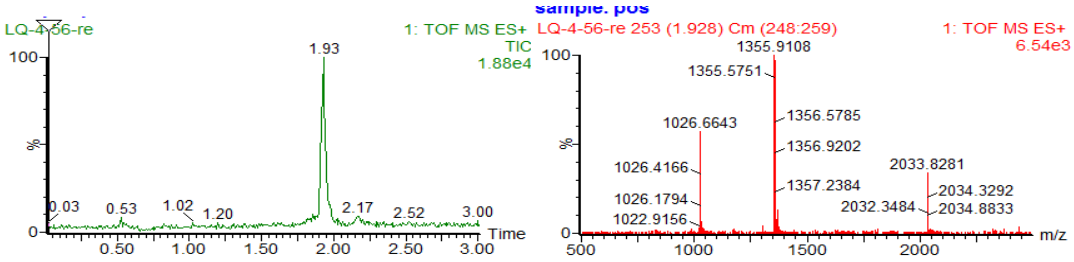

Figure App.1.12. LC-MS trace and ESI-MS of CBM glycoform 12. MS (ESI) Calcd for $C_{171}H_{255}N_{43}O_{64}S_4$: $[M+2H]^{2+}$ m/z = 2032.35, $[M+3H]^{3+}$ m/z = 1355.23, $[M+4H]^{4+}$ m/z = 1016.67.

CBM glycoform 13: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 174.0 mg of the crude glycopeptide was obtained. 3.8 mg (0.795 μmol) of the crude material was dissolved in 200 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 16 mL of folding buffer. After folding and centrifugal

concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **13** (0.32 mg, white solid, 7% yield based on resin loading).

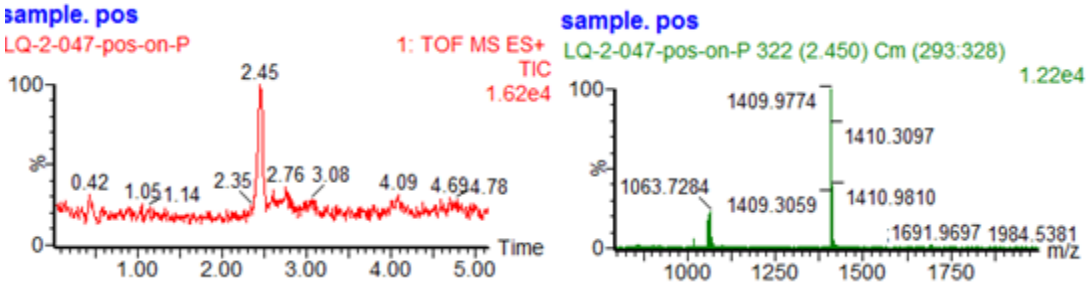

Figure App.1.13. LC-MS trace and ESI-MS of CBM glycoform 13. MS (ESI) Calcd for $C_{177}H_{265}N_{43}O_{69}S_4$: $[M+2H]^{2+}$ m/z = 2113.37, $[M+3H]^{3+}$ m/z = 1409.25, $[M+4H]^{4+}$ m/z = 1057.19.

CBM glycoform 14: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 167.1 mg of the crude glycopeptide was obtained. 8.6 mg (1.705 μmol) of it was dissolved in 450 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 40 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **14** **(**0.43 mg, white solid, 4% yield based on resin loading).

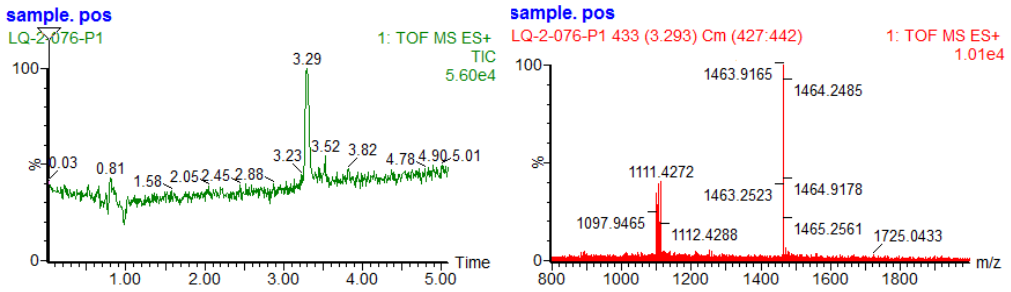

Figure App.1.14. LC-MS trace and ESI-MS of CBM glycoform 14. MS (ESI) Calcd for $C_{183}H_{275}N_{43}O_{74}S_4$: $[M+2H]^{2+}$ m/z = 2194.40, $[M+3H]^{3+}$ m/z = 1463.27, $[M+4H]^{4+}$ m/z = 1097.70.

CBM glycoform 15: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 178.9 mg of the crude glycopeptide was obtained. 8.0 mg (1.705 μmol) of it was stirred in 450 μL of the hydrazine solution at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 40 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **15 (**0.43 mg, white solid, 5% yield based on resin loading).

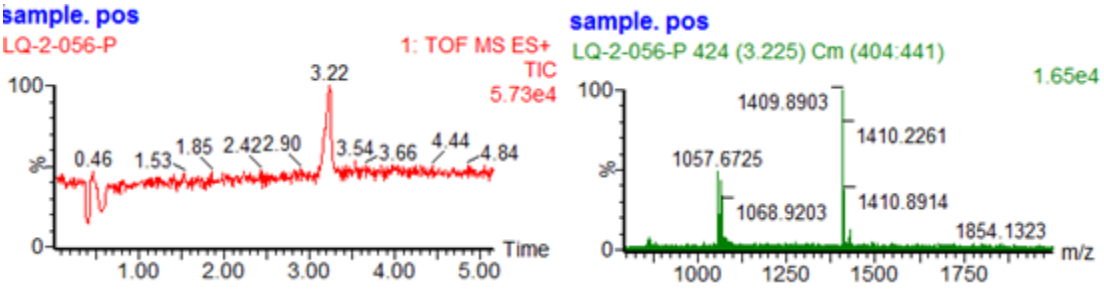

Figure App.1.15. LC-MS trace and ESI-MS of CBM glycoform 15. MS (ESI)Calcd for $C_{177}H_{265}N_{43}O_{69}S_4$: $[M+2H]^{2+}$ m/z = 2113.37, $[M+3H]^{3+}$ m/z = 1409.25, $[M+4H]^{4+}$ m/z = 1057.19.

CBM glycoform 16: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 280 mg of the crude glycopeptide was obtained. 8.0 mg (1.705 μmol) of it was dissolved in 450 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 40 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **16** **(**1.25 mg, white solid, 20% yield based on resin loading).

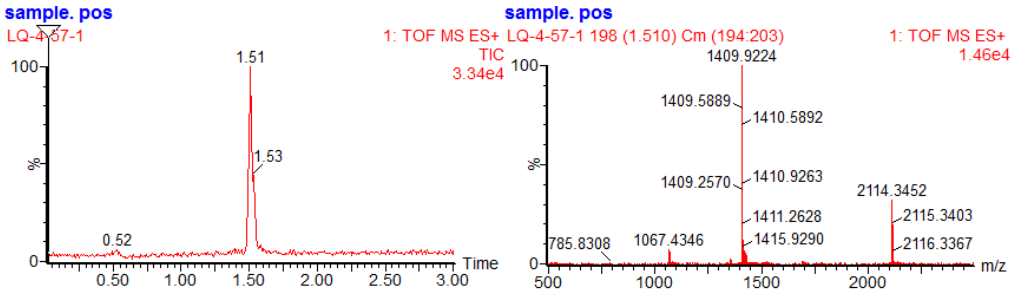

Figure App.1.16. LC-MS trace and ESI-MS of CBM glycoform 16. MS (ESI)Calcd for $C_{177}H_{265}N_{43}O_{69}S_4$: $[M+2H]^{2+}$ m/z = 2113.37, $[M+3H]^{3+}$ m/z = 1409.25, $[M+4H]^{4+}$ m/z = 1057.19.

CBM glycoform 17: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 170.5 mg of the crude glycopeptide was obtained. 8.6 mg (1.705 μmol) of the crude material was dissolved in 450 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 40 mL of folding buffer. After folding and centrifugal

concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **17** (0.40 mg, white solid, 4% yield based on resin loading).

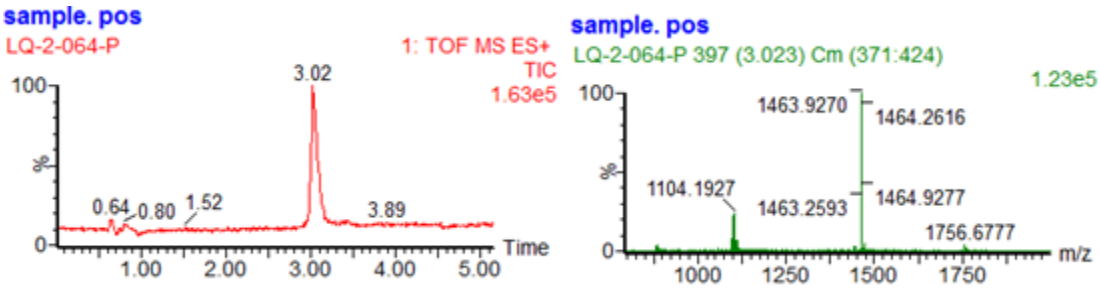

Figure App.1.17. LC-MS trace and ESI-MS of CBM glycoform 17. MS (ESI) Calcd for $C_{183}H_{275}N_{43}O_{74}S_4$: $[M+2H]^{2+}$ m/z = 2194.40, $[M+3H]^{3+}$ m/z = 1463.27, $[M+4H]^{4+}$ m/z = 1097.70.

CBM glycoform 18: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 167.4 mg of the crude glycopeptide was obtained. 9.1 mg (1.705 μmol) of it was dissolved in 450 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 16 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **18** (0.40 mg, white solid, 3% yield based on resin loading).

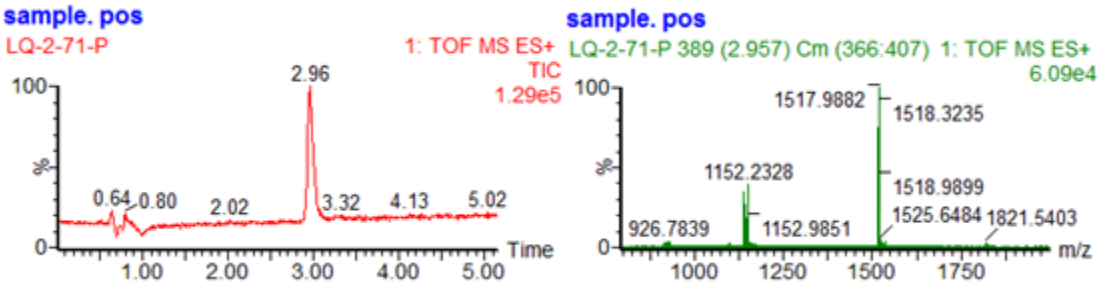

Figure App.1.18. LC-MS trace and ESI-MS of CBM glycoform 18. MS (ESI) Calcd for $C_{189}H_{285}N_{43}O_{79}S_4$: $[M+2H]^{2+}$ m/z = 2275.42, $[M+3H]^{3+}$ m/z = 1517.28, $[M+4H]^{4+}$ m/z = 1138.21.

CBM glycoform 19: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage and lyophilization, 188.6 mg of the crude glycopeptide was obtained. 16 mg (2.86 μmol) of the it was dissolved in 1000 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 16 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **19** (0.77 mg, white solid, 4% yield based on resin loading).

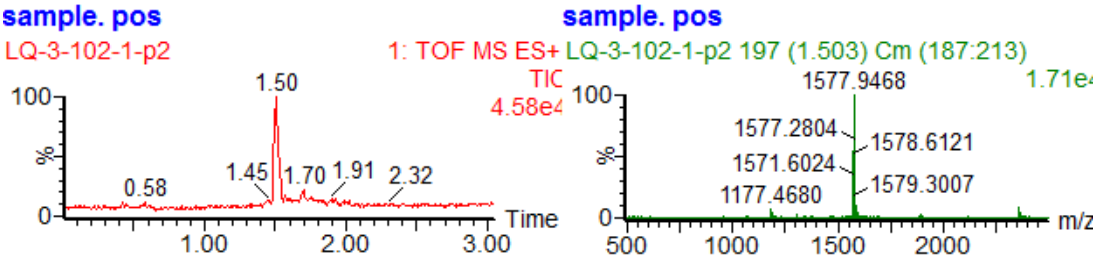

Figure App.1.19. LC-MS trace and ESI-MS of CBM glycoform 19. MS (ESI) Calcd for $C_{195}H_{295}N_{43}O_{84}S_4$: $[M+2H]^{2+}$ m/z = 2356.46, $[M+3H]^{3+}$ m/z = 1571.30, $[M+4H]^{4+}$ m/z = 1178.73.

CBM glycoform 20: The glycosylated peptide was synthesized on a 0.05 mmol scale. After cleavage, 165.0 mg of the crude glycopeptide was obtained. 16 mg (2.48 μmol) of it was dissolved in 1000 μL of the hydrazine solution and stirred at room temperature for 30 min under helium. The reaction was quenched with the acetic acid solution and the pH was adjusted to between 8 and 9. The resulting glycopeptide solution was then diluted by the addition of 16 mL of folding buffer. After folding and centrifugal concentration, HPLC purification with a linear gradient of 20→40% acetonitrile in H_2O over 30 min afforded the correctly folded **20** (0.74 mg, white solid, 3% yield based on resin loading).

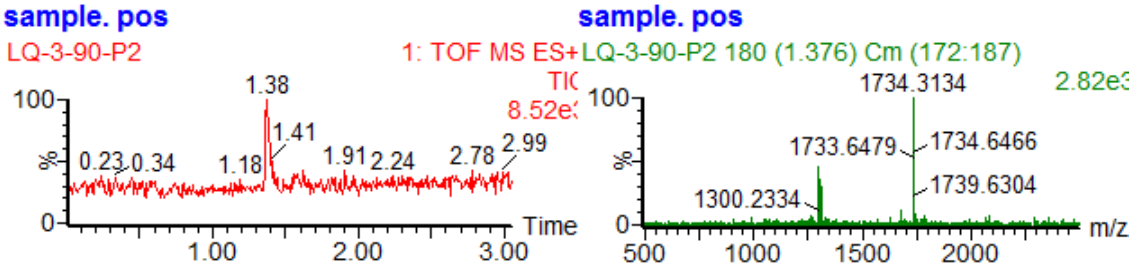

Figure App.1.20. LC-MS trace and ESI-MS of CBM glycoform 20. MS (ESI) Calcd for $C_{213}H_{325}N_{43}O_{99}S_4$: $[M+2H]^{2+}$ m/z = 2599.53, $[M+3H]^{3+}$ m/z = 1733.35, $[M+4H]^{4+}$ m/z = 1300.27.

App.2. The Stability and Binding Data for the 20 CBM Glycoforms with Uncertainties

Table App.2.1. Stability and binding data for the 20 CBM glycoforms with uncertainties. Half-Life to Thermolysin Degradation and T_m results are presented as mean of three trials ± SD. Adsorption affinity consant, K_{ads} and B_{max} results are presented as the mean of two trials ± SD. *Denotes an averaged value of four trials ± SD. **Weak affinity to cellulose noted, no K_{ads} value could be obtained.

CBM Glycoform	Half-Life to Thermolysin Degradation (hr)	T_m (°C)	K_{ads} (μM^{-1})	B_{max} (μmol/g)
1	0.23 ± 0.02	62.2 ± 0.6	0.0894 ± 0.0007*	24 ± 5*
2	0.28 ± 0.02	61.1 ± 0.7	0.16 ± 0.02	6.5 ± 0.7
3	0.23 ± 0.02	59.8 ± 0.6	0.34 ± 0.16	25 ± 7
4	0.208 ± 0.002	65 ± 1	0.06 ± 0.03	22 ± 11
5	1.09 ± 0.01	70.1 ± 0.9	0.4 ± 0.2	6 ± 1.3
6	2.13 ± 0.06	73.2 ± 0.7	0.21 ± 0.05	3.6 ± 0.8
7	2.10 ± 0.05	71.7 ± 0.8	0.13 ± 0.02	3 ± 1.0
8	0.49 ± 0.01	67.3 ± 0.6	0.35 ± 0.17	5 ± 2
9	0.55 ± 0.01	65.5 ± 0.7	0.31 ± 0.08	5.9 ± 0.7
10	0.54 ± 0.01	64.5 ± 0.5	0.25 ± 0.02	10.5 ± 0.6
11	1.82 ± 0.04	74 ± 1	0.19 ± 0.09	13 ± 6
12	1.96 ± 0.07	72 ± 1	0.268 ± 0.002	16 ± 1.2
13	4.33 ± 0	75 ± 2	0.66 ± 0.05	6.6 ± 0.5
14	10.5 ± 0.8	77.9 ± 0.6	0.373 ± 0.008	9.6 ± 0.11
15	2.8 ± 0.2	75.4 ± 0.2	0.245 ± 0.003	5.6 ± 0.18
16	3.6 ± 0.4	75.2 ± 0.5	0.19 ± 0.02	7.1 ± 0.9
17	9.1 ± 0.6	77.4 ± 0.3	0.58 ± 0.08	4.9 ± 0.15
18	9.1 ± 0.6	76 ± 2	0.22 ± 0.01	13.4 ± 0.4
19	3.36 ± 0.08	77 ± 2	0.155 ± 0.012	8.1 ± 0.5
20	5.5 ± 0.4	75.1 ± 0.8	~0**	~0**

App.3. The BMCC Binding Curves for the 20 CBM Glycoforms

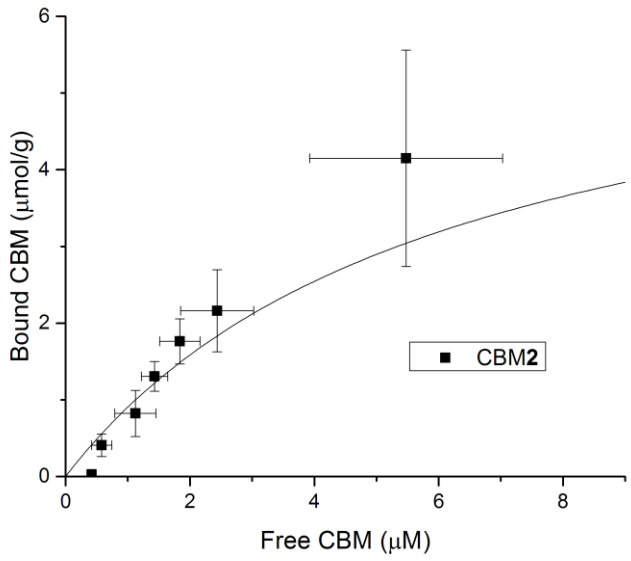

Figure App.3.1. The binding curve for CBM glycoform 2. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

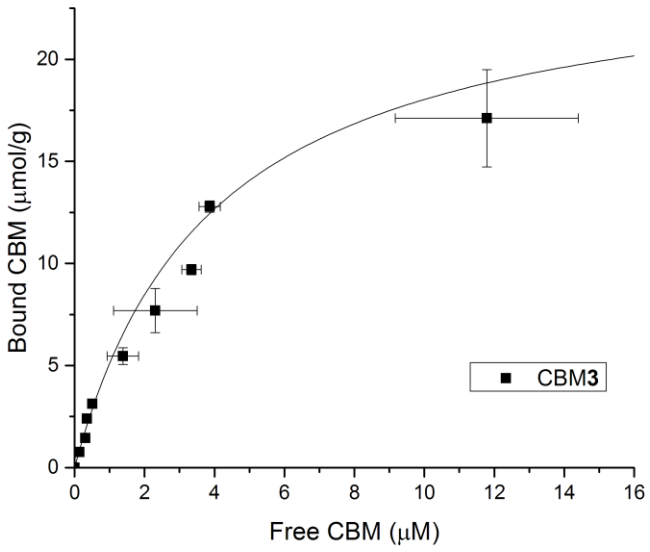

Figure App.3.2. The binding curve for CBM glycoform 2. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

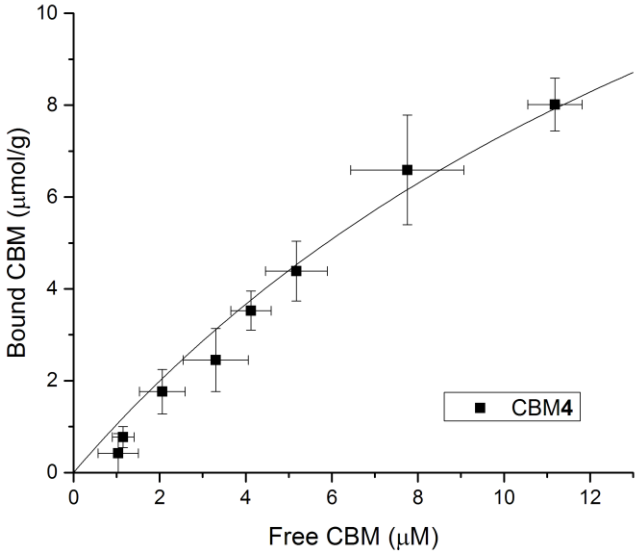

Figure App.3.3. The binding curve for CBM glycoform 4. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

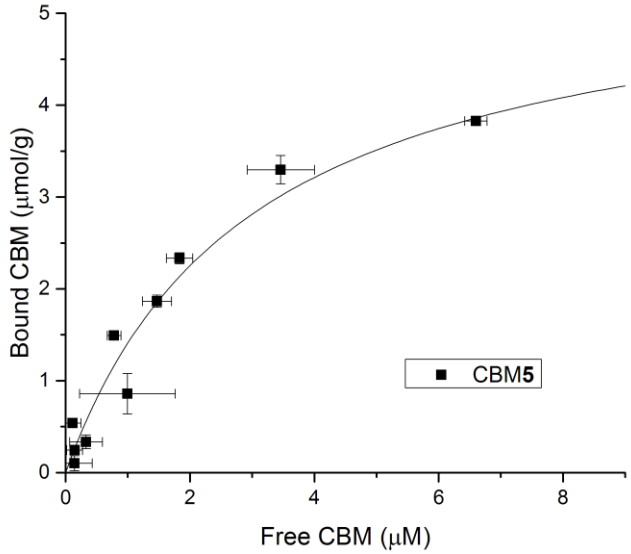

Figure App.3.4. The binding curve for CBM glycoform 5. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

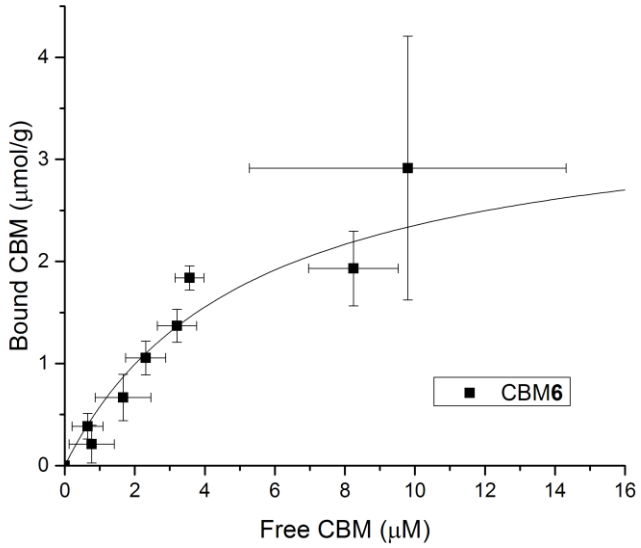

Figure App.3.5. The binding curve for CBM glycoform 6. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

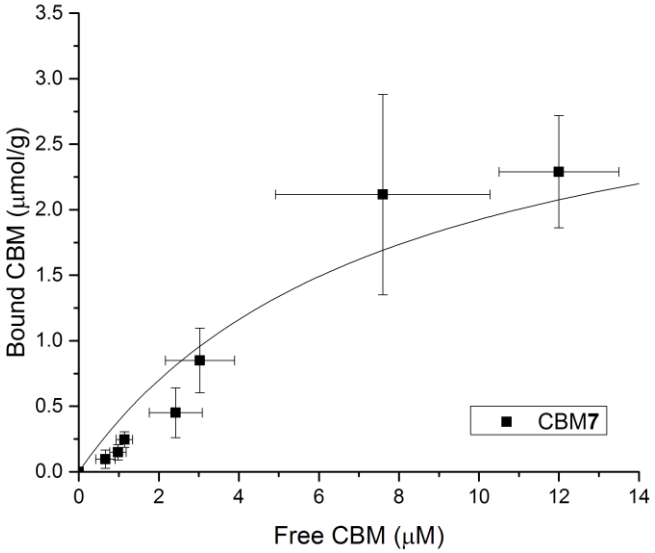

Figure App.3.6. The binding curve for CBM glycoform 7. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

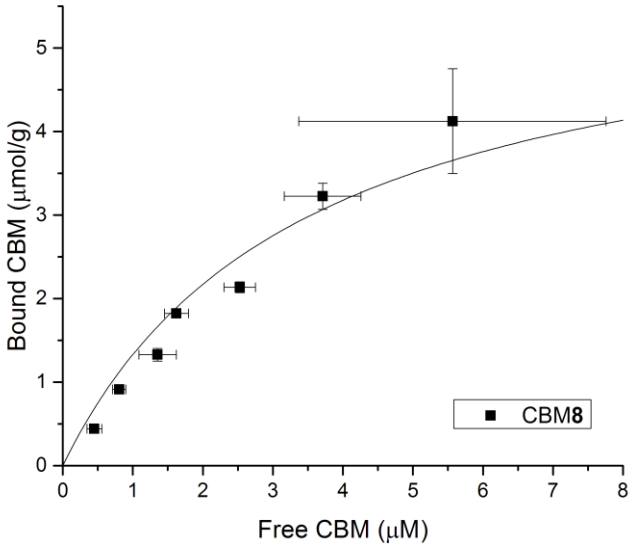

Figure App.3.7. The binding curve for CBM glycoform 8. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

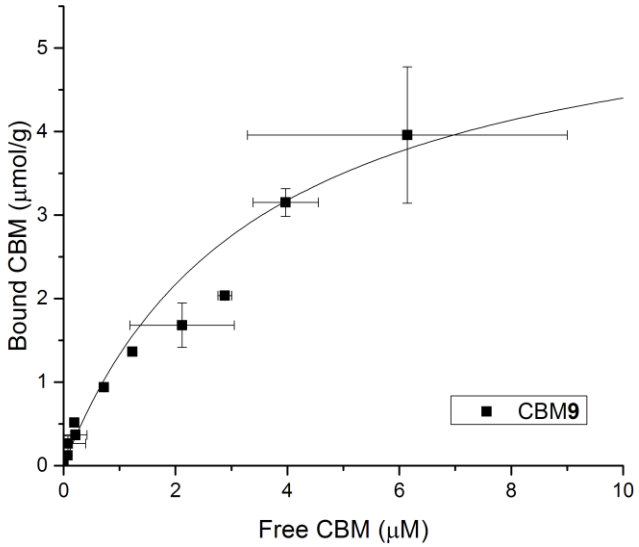

Figure App.3.8. The binding curve for CBM glycoform 9. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

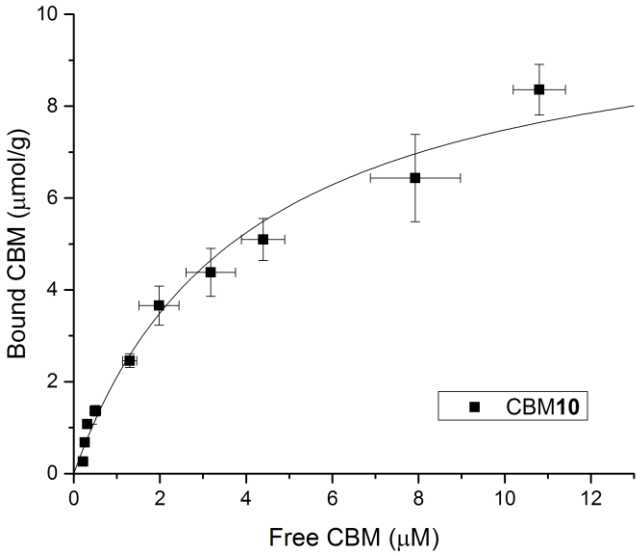

Figure App.3.9. The binding curve for CBM glycoform 10. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

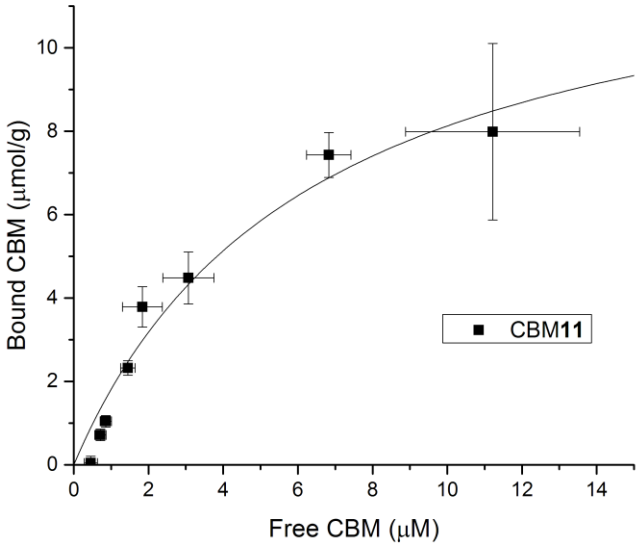

Figure App.3.10. The binding curve for CBM glycoform 11. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

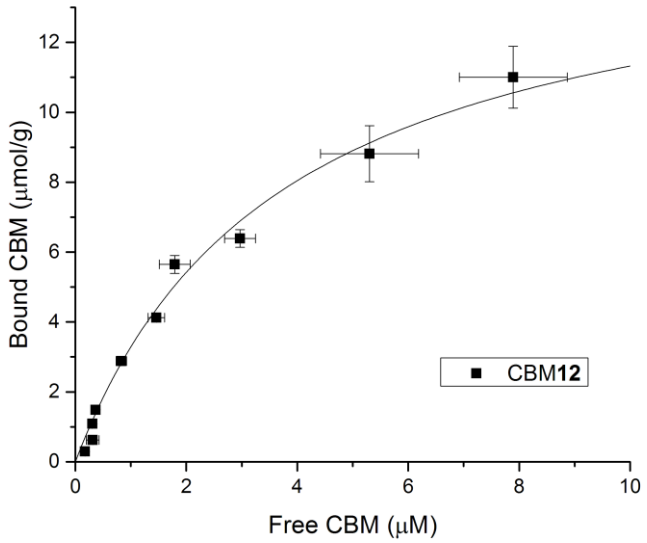

Figure App.3.11. The binding curve for CBM glycoform 12. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

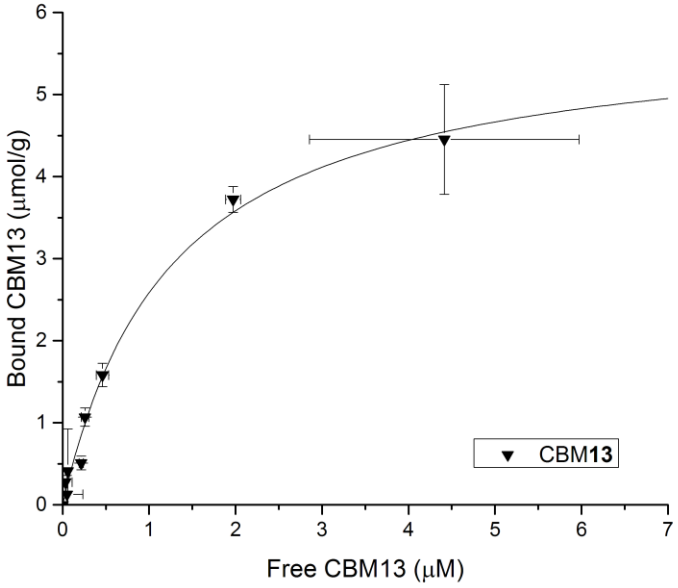

Figure App.3.12. The binding curve for CBM glycoform 13. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

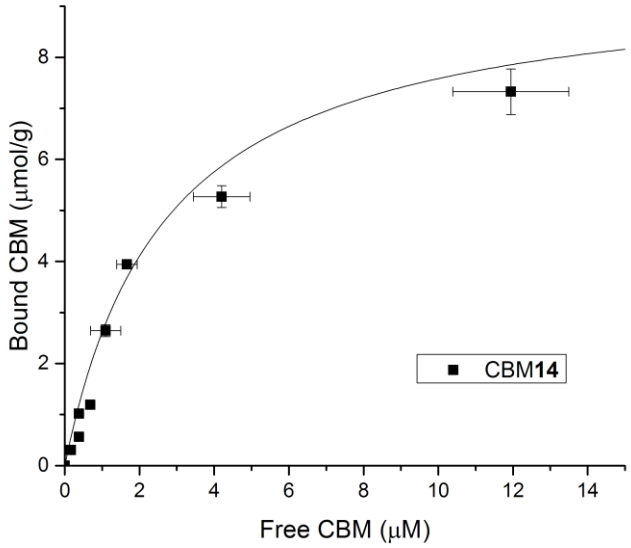

Figure App.3.13. The binding curve for CBM glycoform 14. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

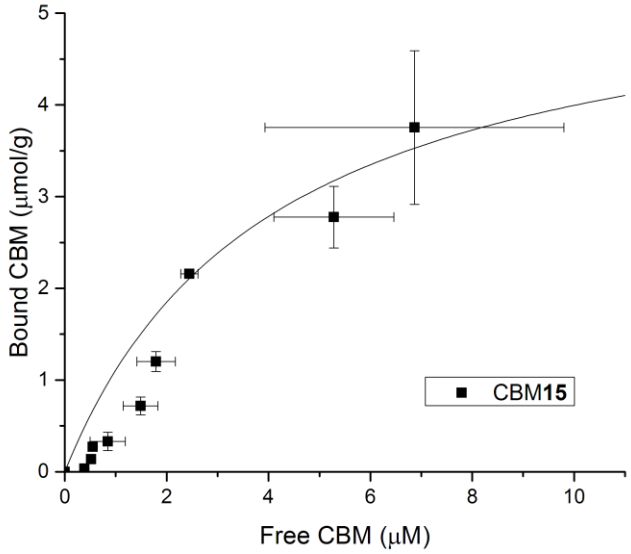

Figure App.3.14. The binding curve for CBM glycoform 15. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

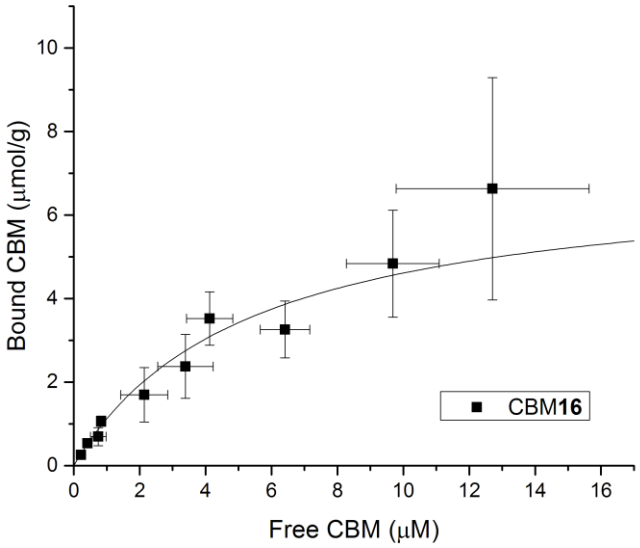

Figure App.3.15. The binding curve for CBM glycoform 16. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

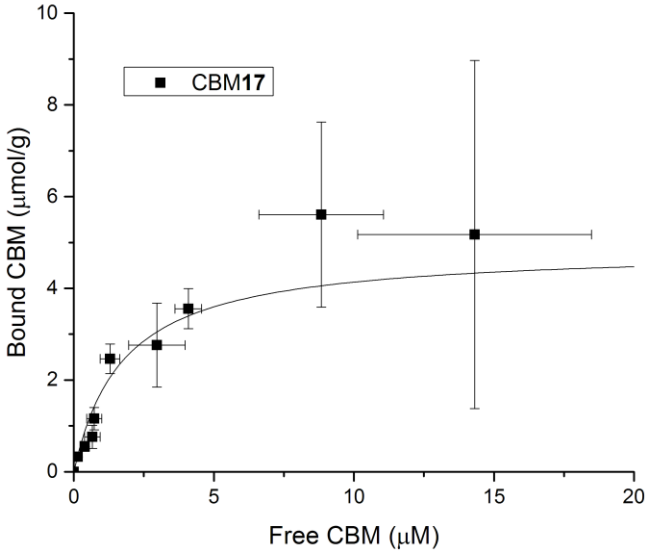

Figure App.3.16. The binding curve for CBM glycoform 17. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

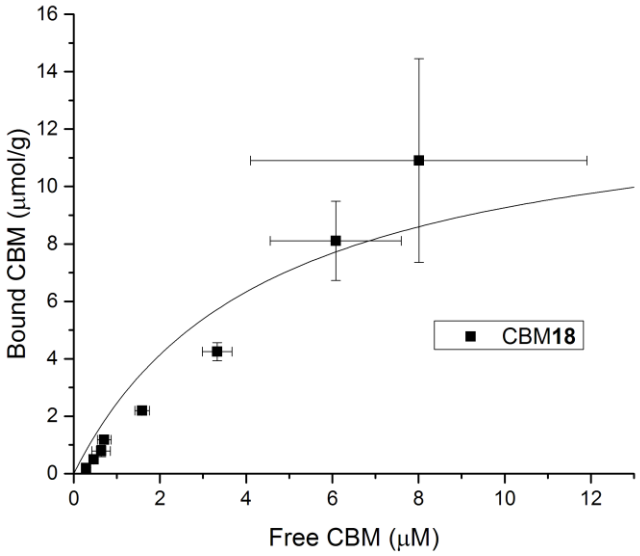

Figure App.3.17. The binding curve for CBM glycoform 18. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

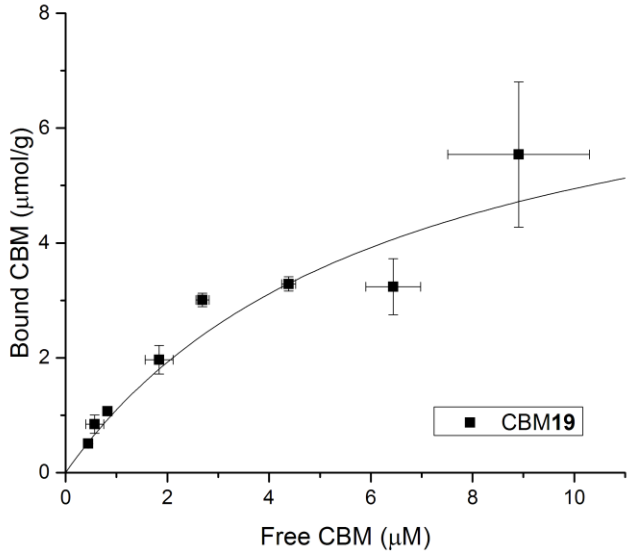

Figure App.3.18. The binding curve for CBM glycoform 19. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.

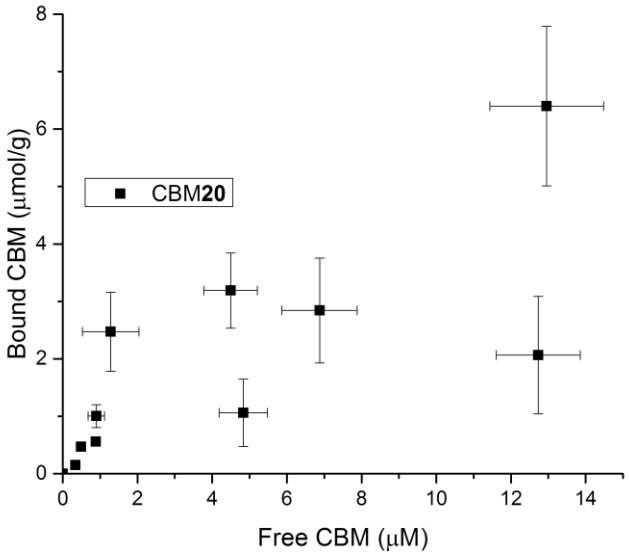

Figure App.3.19. The binding curve for CBM glycoform 20. Data were fit using Equation 4.2 and binding constants are provided in Table App.4.1.